

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:03:27 ; Search time 10.6437 Seconds  
(without alignments)  
2330.267 Million cell updates/sec

Title: US-09-813-453A-49  
Perfect score: 1321  
Sequence: 1 MIFVLDVGNTNTVLGVYDGD.....PFLTLTGLKLLYEKTEKKG 258

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	975	73.8	254	2 F83660	hypothetical prote
2	900	68.1	233	2 S66100	conserved hypothet
3	855.5	64.8	259	2 AF1102	conserved hypothet
4	848.5	64.2	259	2 AF1464	conserved hypothet
5	755	57.2	273	2 E97293	probable transcrip
6	663	50.2	265	2 T36391	hypothetical prote
7	575	43.5	261	2 B87489	transcription acti
8	504.5	38.2	274	2 H86937	conserved hypothet
9	479.5	36.3	272	2 A70955	hypothetical prote
10	415.5	31.5	262	2 E75516	conserved hypothet
11	375.5	28.4	246	2 D72320	conserved hypothet
12	342.5	25.9	273	2 D71326	conserved hypothet
13	256.5	19.4	262	2 F70165	conserved hypothet
14	150	11.4	257	2 S75559	hypothetical prote
15	148.5	11.2	229	2 E70465	hypothetical prote
16	148	11.2	276	2 AI2292	hypothetical prote
17	146	11.1	592	2 B81009	Bira protein/Bvg a
18	146	11.1	592	2 H82031	probable biotin-la
19	140.5	10.6	267	2 I40327	baf protein - Bord
20	123.5	9.3	248	2 H83111	hypothetical prote
21	115	8.7	242	2 A82637	conserved hypothet
22	105.5	8.0	597	2 B69251	probable electron
23	105	7.9	224	2 A99571	conserved hypothet
24	103.5	7.8	512	2 A98352	hydantoinase homol
25	103.5	7.8	512	2 AE2930	hydantoinase A [im
26	99.5	7.5	223	2 G71887	hypothetical prote
27	97	7.3	434	2 E96961	membrane protein c
28	96	7.3	394	2 AB1068	probable membrane
29	95	7.2	5825	2 TL2117	polyprotein - fava

30	94	7.1	153	2 A95928	probable amine oxi
31	94	7.1	467	2 AE3142	hypothetical prote
32	94	7.1	520	2 G98145	probable aminotran
33	93.5	7.1	443	2 E88343	protein Y38FlA.6 l
34	93.5	7.1	465	2 T26686	hypothetical prote
35	93	7.0	422	2 AD0484	GntP family permea
36	88.5	6.7	818	2 T15803	hypothetical prote
37	88	6.7	425	2 T43861	glutamate-1-semial
38	87.5	6.6	223	2 F64627	hypothetical prote
39	87	6.6	473	2 C64371	aspartate kinase (
40	86.5	6.5	578	2 T03475	transmembrane prot
41	86	6.5	267	2 D70819	cyclase hisF [slmi
42	85	6.4	323	2 A70324	hydroxymethylbilan
43	85	6.4	503	2 D70930	hypothetical prote
44	84.5	6.4	366	2 AH0939	probable glycerol
45	84.5	6.4	758	2 T39210	3-isopropylmalate

ALIGNMENTS

RESULT 1  
F83660  
hypothetical protein BH0086 [imported] - Bacillus halodurans (strain C-125)  
C;Species: Bacillus halodurans  
C;Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C;Accession: F83660  
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.;  
Nucleic Acids Res. 28, 4317-4331, 2000  
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans  
A;Reference number: A83650; MUID:20512582; PMID:11058132  
A;Accession: F83660  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-254 <STO>  
A;Cross-references: GB:AP001507; GB:BA000004; NID:gl0172612; PIDN:BAB03805.1; GSPDB:  
A;Experimental source: strain C-125  
C;Genetics:  
A;Gene: BH0086  
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 73.8%; Score 975; DB 2; Length 254;  
Best Local Similarity 70.8%; Pred. NO. 9.6e-78;  
Matches 179; Conservative 36; Mismatches 38; Indels 0; Gaps 0;

QY	1	MIFVLDVGNTNTVLGVYDGD	DELKHHWRIETSRKTEDEYGMKALLNHVGLQFSDIRGI	60
Db	1	MILVIDVGNTNTVLGVYQDET	LVHHWRLATSRQKTEDEYAMTVRSFLDHAGLQFQDIDGI	60
QY	61	IISSVVPPIPMFALERMCLK	FYHIKPLIVGPGIKTGLDIKIDYDNPREVGADRIVNAVAGIHL	120
Db	61	VISSVVPPIPMFSLQMKCKK	FYHVTPTMIIGPGIKTGLNIKIDYDNPREVGADRIVNAVAAIEL	120
QY	121	YGSPLIIVDFGTATTTCYIN	EHKQYMGGAIAPIAGIMISTEALFARAAKLPRIEIAIPDDII	180
Db	121	YGYPPIVDFGTATTTCYIN	EKKQYAGGVIAPIAGIMISTEALYHRASKLPRIEIAIPKQVV	180
QY	181	GKNTVSAMQAGILYGVQVE	GIVSRMKAKSKIPPKVIATGGLAPLIASESIIIDVDPF	240
Db	181	GTNTIDSMQSGIFYGVSVQ	VDGVVVKRMAQAESEPKVIATGGLAKLIGTESETIDVDSF	240
QY	241	LTLTGLKLLYEKN	253	
Db	241	LTLKGLQLIYKKN	253	

RESULT 2  
S66100  
conserved hypothetical protein yacB - Bacillus subtilis  
C;Species: Bacillus subtilis  
C;Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 18-Aug-2000  
C;Accession: S66100; E69740  
R;Ogasawara, N.; Nakai, S.; Yoshikawa, H.









C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C;Accession: D72320  
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;  
C.M.  
Nature 399, 323-329, 1999  
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A;Reference number: A72200; MUID:99287316; PMID:10360571  
A;Accession: D72320  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-246 <ARN>  
A;Cross-references: GB:AE001754; GB:AE000512; NID:g4981417; PIDN:AAD35964.1; PID:g498141  
A;Experimental source: strain MSB8  
C;Genetics:  
A;Gene: TM0883  
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match	28.4%	Score 375.5;	DB 2;	Length 246;
Best Local Similarity	37.5%;	Pred. No. 2.4e-25;		
Matches 94;	Conservative 48;	Mismatches 92;	Indels 17;	Gaps 9;
QY	1	MIFVLVDGNTNTVLGVYDGDDELKHHWRIETSRSKTEDEYGMKIKALLNHVGLQFSDIRGI	60	
Db	1	MYLLVDVGNTHSVFSITEDGKTFRRWLSTGVFQTEDELSHLHPLL---	GDAMREIKGI	57
QY	61	IISSVVPPIMFALERMCLKYFHIKPLIVGPGIKTGLDIKYD--NPREVGADRIVNAVAGI	118	
Db	58	GVASVVPQTQNTVIERFSQKYFHSPIWV--KAKNGC-VKWNVKNPSEVGADRIVNAVAFV	114	
QY	119	HLYGSPLIIVDFGTATTYCYINEHKQYMGGAIAPIGIMISTEALFARAAKLPRIEIARPDD	178	
Db	115	KEYGKNGLIIDMGATATVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKLPLVEVKPADF	173	
QY	179	IIGKNTVSAMQAGILYGVGVQVEGIVSRMK-AKSKIPPKVATGGLAPLIASESDII--D	235	
Db	174	VVGKDTEENIRLGVNGSVYALEGIIIGRIKEVYGDLDP--VVLTGGQSKIV---KDMIKHE	228	
QY	236	VDPDFLTLTGL	246	
Db	229	IFEDELTIKGV	239	

RESULT 12  
D71326  
conserved hypothetical protein TP0431 - syphilis spirochete  
-C; Species: Treponema pallidum subsp. pallidum (syphilis spirochete)  
; Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 20-Jun-2000  
; Accession: D71326  
R: Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin-  
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDo-  
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
Science 281, 375-388, 1998  
A: Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.  
A: Reference number: A71250; MUID: 98332770; PMID: 9665876  
A: Accession: D71326  
A: Status: preliminary; nucleic acid sequence not shown; translation not shown  
A: Molecule type: DNA  
A: Residues: 1-273 <COL>  
A: Cross-references: GB:AE001220; GB:AE000520; NID: g3322705; PIDN: AAC65417.1; PID: g332271  
A: Experimental source: strain Nichols  
C: Genetics:  
A: Gene: TP0431  
C: Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Qy	2 IFVLGVGNTNIVLGVYDGE-----LKHHWRJETSRSKTEDEYEGMMIKALLNHVCLQFSD 56 : : : : : :     : : : :             : : : :     :
Dd	1 MLLIDVGNSHVFGI -QGNGGRVCVREFLRLAPDARKTQDEYSLLIHALCERAGVGRAS 59 : : : : : :     : : : :             : : : :     :

QY	57	IRGIIISSVPPIMFALERMCLKYFFHIKPLIVGPGIKTGLDIKYDNP--REVGDRIVNA	114
Dd	60	LRFDAFISSVVPVLTKTIADAVAQISGVQPWFVGWPAYEHLPVRIPEPVRAEIGTDLVANA	119
QY	115	VAGIHLXGSPLIIIVDFGTATTCYCYNIEHKQYMGGAIAPGIMISTEALFAARAKLPRIEIA	174
Dd	120	VAAYVHFSACVVDCGTALTFTAVDGTGLIQGVAIAPGLRTAVQSLHTGTQAQLPLVPLA	179
QY	175	RPDDIIGKNTVSAMQAGILGYVQGVEGIVSRMKAKSKIIPPKVIATCGGLAPLIASESDII	234
Dd	180	LPDSVLGKDTHHAVQAGVVRGTLFVIRAMIAOCCQKELGCRCAAVITGGLSRLFSSEVD-F	238
QY	235	DVDPFLTLTGL	246
Dd	239	PPIDAQLTLSGL	250

RESULT 13  
F70165  
conserved hypothetical protein BB0527 - Lyme disease spirochete  
C:Species: Borrelia burgdorferi (Lyme disease spirochete)  
C:Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 08-Oct-1999  
C:Accession: F70165  
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathig  
son, D.; Peterson, J.; Kervage, A.R.; Quackenbush, J.; Salzberg, S.; Hans  
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.;  
Nature 390, 580-586, 1997  
A:Authors: Smith, H.O.; Venter, J.C.  
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi  
A:Reference number: A70100; MUID:98065943; PMID:9403685  
A:Accession: F70165  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-262 <KLE>  
A:Cross-references: GB:AE001154; GB:AE000783; NID:92688431; PIDN:AAC66882.1  
A:Experimental source: strain B31

Query Match	19.4%;	Score 256.5;	DB 2;	Length 262;
Best Local Similarity	26.3%;	Pred. No. 6.6e-15;		
Matches 67;	Conservative 59;	Mismatches 108;	Indels 21;	Gaps 5;

QY	4	VLDVGNTNTVLGVYDGDDELKHHWRIETSRSKTEDEYGMIMKALLNHVGLQFSDIRGIIS	63
Dd	10	IIDIGNTSIAFALEFDNQVNLFIKMKTNMLRYDEVYSFFEEFNDF-----NVNKVFIS	63
QY	64	SVVPPIMFALERMCLKYFHFIKPLIVPGIKTGLDIKYD---NPRE-----VGADRIVNAV	115
Dd	64	SWPILNETFKNVIFSPFKIKPLFI-----GFDLNYDLTFNPKSDKFLLGSDVFANLV	117
QY	116	AGIHLYG-SPLIIVDFTATTYYCINEHKQYMGGAIAPGIMISTEALFARAAKLPRIEIA	174
Dd	118	AAIENYSFENVLVDLTACTIFAVSRQDGILGGIINSGPLINFNSLLDNAYLIKKEPIS	177
QY	175	RPDDIIGKNTVSMAOAGILYGYVQVEGIVSRMKAKSKIPPKVIATGGLAPLIASESDII	234
Dd	178	TPNNLLERTTGSVNSGLFYQYKYLIEGVYRDIQMYYKKKNLIITGGNADLILSLIEIE	237
QY	235	DVVDPFLLTLGLKL	249
Dd	238	FIFNIHILTVEGVRIIL	252

RESULT 14  
S75559  
hypothetical protein slr0812 - *Synechocystis* sp. (strain PCC 6803)  
C:Species: *Synechocystis* sp.  
A:Variety: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999  
C:Accession: S75559  
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.;  
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamad  
DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium



•  
52

C;Genetics:  
A;Start codon: GTG

Matches 41; Conservative 30; Mismatches 39; Indels 14; Gaps 5;

Db 100 GIDR---ALAGLGTGLTYGFPCLVVDGGTALTITGFDQDKKLVGGAILPGIGLQLATLGD 156

157 RLAA LPKLEMDQLTELPDRWALDTPSAIFSGVVYGVLGALQ---SYLQDWOKLFPGAAMV 213

Dbb 214 ITGG 217

R; Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov

**A; Accession: E70465**

A:Gene: ag 1924

Matches	Conservative	Mismatches	Indels	Gaps
60	43	111	37	9

Db 6 VDVGNSSVDIALWEGKKVKDF--LKL SHEEFFLKEEFFPKLKALGISVKOSFSEKVRG----

60 -KIPKIKFLKKE-----NPIOVDPYKTPETLGTDRVALAYSACKFEYVK 101

```
db 102 NVVVISAGTALVIDLVLEK - FKGGFITLGLGKKLKILSDLAEGIPPEFFPEEVEIFLGRS 160
```

```

db      161  TRECVLGG---AYRESTEEIKSTLKLWRKVEKRFKFWTTGGEGKYFSK-----FGIYDPL 213

```

SECRET

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:51:40 ; Search time 5.18991 Seconds  
(without alignments)  
2061.866 Million cell updates/sec

Title: US-09-813-453A-49  
Perfect score: 1321  
Sequence: 1 MIFVLDVGNTNTVLGVYDGD.....PFLTTLTGLKLLYEKNTKKG 258

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892.seqs, 41476328 residues  
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	900	68.1	233	1 YACB_BACSU	P37564 bacillus su
2	140.5	10.6	267	1 BAF_BORPE	Q45338 bordetella
3	87	6.6	425	1 GSA_CLOPE	Q9znc8 clostridium
4	87	6.6	473	1 AK_METJA	Q57991 methanococc
5	86	6.5	267	1 HIS6_MYCTU	Q53908 mycobacteri
6	86	6.5	303	1 V212_FOWPV	Q9J523 fowlpox vir
7	85	6.4	304	1 HEM3_AQUAE	Q66621 aquifex ae
8	84.5	6.4	758	1 LEU2_SCHPO	O14289 schizosacch
9	84	6.4	361	1 OP22_HAEIN	P20149 haemophilus
10	84	6.4	419	1 Y092_HAEIN	Q57493 haemophilus
11	83.5	6.3	469	1 LEU2_RHILO	Q98ef1 rhizobium l
12	83	6.3	365	1 HHP1_SCHPO	P40235 schizosacch
13	83	6.3	401	1 CHRA_ALCEU	P17551 alcaligenes
14	83	6.3	491	1 CIMA_METJA	Q58787 methanococc
15	82.5	6.2	420	1 YD43_METJA	Q58739 methanococc
16	82	6.2	336	1 GLPX_ECOLI	P28860 escherichia
17	81.5	6.2	258	1 HIS6_CORGL	Q31139 corynebacte
18	81.5	6.2	429	1 AROA_METJA	Q57925 methanococc
19	81.5	6.2	469	1 LEU2_AGR5	Q8uby9 agrobacteri
20	81.5	6.2	494	1 HR25_YEAST	P29295 saccharomyc
21	81	6.1	424	1 SAHH_METJA	P58855 methanopyru
22	81	6.1	543	1 CH60_CLOAB	P30717 clostridium
23	81	6.1	610	1 FTSL_HAEIN	P45059 haemophilus
24	81	6.1	1184	1 POL2_GFLV	P18474 grapevine f
25	80.5	6.1	470	1 LEU2_AZOTI	P96195 azotobacter
26	80.5	6.1	579	1 YD49_MYCTU	Q11019 mycobacteri
27	80	6.1	359	1 OP21_HAEIN	P43839 haemophilus
28	80	6.1	788	1 PUR2_YARLI	Q99148 y bifunctio
29	79.5	6.0	469	1 LEU2_RHIME	Q92176 rhizobium m
30	79	6.0	1042	1 CARB_HA1N1	Q9hp43 halobacteri
31	79	6.0	4451	1 GR5B_BACBR	P14688 b gramicidi
32	78.5	5.9	232	1 UPP_PYRAB	Q9v0k1 pyrococcus
33	78.5	5.9	410	1 RAA2_CHLRE	Q9smh4 chlamydomon

RESULT 1				
YACB_BACSU				
ID	YACB_BACSU	STANDARD;	PRT;	233 AA.
AC	P37564;			
DT	01-OCT-1994 (Rel. 30, Created)			
DT	01-OCT-1994 (Rel. 30, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Hypothetical protein yacB.			
GN	YACB.			
OS	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			
OX	NCBI_TaxID=1423;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RX	MEDLINE=96051385; PubMed=7584024;			
RA	Ogasawara N., Nakai S., Yoshikawa H.;			
RT	"Systematic sequencing of the 180 kilobase region of the Bacillus subtilis chromosome containing the replication origin.";			
RL	DNA Res. 1:1-14(1994).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RX	MEDLINE=98044033; PubMed=9384377;			
RA	Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G., Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S., Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S., Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M., Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A., Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T., Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D., Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N., Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G., Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A., Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L., Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C., Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M., Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C., Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H., Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M., Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y., Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F., Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B., Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K., Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A., Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A., Viari A., Wambutt R., Wedler E., Wedler H., Weltzenegger T., Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K., Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;			
RT	"The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.";			
RL	Nature 390:249-256(1997).			
CC	-I- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.			

ALIGNMENTS

34	73.5	5.9	478	1	GLN1_MYCTU	Q10377 mycobacteri
35	73.5	5.9	485	1	SAHH_STRCO	Q9kzml streptomyc
36	73.5	5.9	543	1	CH60_BARBA	P35635 bartonella
37	73.5	5.9	859	1	YD48_MYCTU	Q11018 mycobacteri
38	78	5.9	296	1	HIS1_YARLI	Q99145 yarrowia li
39	78	5.9	360	1	OP26_HAEIN	Q48216 haemophilus
40	78	5.9	585	1	CH60_PYRSA	P46224 pyrenomonas
41	78	5.9	1146	1	MMLC_MYCTU	Q50585 mycobacteri
42	77.5	5.9	261	1	HIS6_MYCLE	Q9x7c2 mycobacteri
43	77.5	5.9	439	1	MAO1_BACSU	P54572 bacillus su
44	77.5	5.9	469	1	LEU2_BRUME	Q8yjc9 brucella me
45	77.5	5.9	502	1	HXK2_ARATH	P93834 arabidopsis















RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
RD.#;  
RL Science 269:496-512(1995).  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U32694; AAC21770.1; -.  
DR TIGR; HI0092; -.  
DR InterPro; IPR003474; Gntp\_permease.  
DR Pfam; PF02447; Gntp\_permease; 1.  
W Hypothetical protein; Transmembrane; Complete proteome.  
I TRANSMEM 1 21 POTENTIAL.  
T TRANSMEM 24 44 POTENTIAL.  
FT TRANSMEM 66 86 POTENTIAL.  
FT TRANSMEM 101 121 POTENTIAL.  
FT TRANSMEM 174 194 POTENTIAL.  
FT TRANSMEM 216 236 POTENTIAL.  
FT TRANSMEM 242 262 POTENTIAL.  
FT TRANSMEM 280 300 POTENTIAL.  
FT TRANSMEM 311 331 POTENTIAL.  
FT TRANSMEM 349 369 POTENTIAL.  
FT TRANSMEM 396 416 POTENTIAL.  
SQ SEQUENCE 419 AA; 42781 MW; 6DCA54BB0A55FF73 CRC64;

Query Match 6.4%; Score 84; DB 1; Length 419;  
Best Local Similarity 26.6%; Pred. No. 7.1;  
Matches 69; Conservative 34; Mismatches 94; Indels 62; Gaps 16;  
QY 29 ETSRSKTEDEYGMKAL- - -LNHVGLQFSDIRGIISVVPPIMFALERMC-LKYFHI 83  
Db 87 ETITNKLGETRALLALATMILTAVGV-FVDVAVITVS- - -PIALALSRRSLSKAAI 141  
QY 84 KPLIVGPGIKTGLDIKYNPREVGADRIVNAVAGIHLGSPLIIVD- - -FGTATT 135  
Db 142 LLAMIGGG-KAG-NINSPN- - - - -NAIAADTFHLPLTSVMMAGIIPALFGLILT 190  
QY 136 YC- - -YINEHKQVMGGAIPGIMISTEAL- - -FARAAKLPRIE- - - - -ARPDDIIGKNTV 185  
Db 191 YFLAKRLINKGSKYTDKEV- - -IVLETONLPSELTALVAPLVAILLALRPLFDIKVDPL 247  
QY 186 SAMQAGILYG- - -YVQVEGIVS-RMKAKSKIPP- - -KVIAATGGLAPLIASESDIIDVDP 239  
Db 248 IALPLGGLIGAFCMGKLRNINSYAINGLSKMTVPVAIMLLGTGALAGIIAN- - - - - 297  
QY 240 FLTLTGLKLLYEKNTEKKG 258  
Db 298 - - - - -SGLKEVLIQGLEHSG 312

RESULT 11  
LEU2\_RHILO  
ID LEU2\_RHILO STANDARD; PRT; 469 AA.  
AC Q98EF1;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE 3-Isopropylmalate dehydratase large subunit (EC 4.2.1.33)  
DE (Isopropylmalate isomerase) (Alpha-IPM isomerase) (IPMI).  
GN LEUC OR MLL4272.  
OS Rhizobium loti (Mesorhizobium loti).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Phyllobacteriaceae; Mesorhizobium.  
OX NCBI\_TaxID=381;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN-MAFF303099;  
RX MEDLINE=21082930; PubMed=11214968;  
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,  
RA Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
RT Mesorhizobium loti.";  
RL DNA Res. 7:331-338(2000).  
CC -1- FUNCTION: Catalyzes the isomerization between 2-isopropylmalate  
CC and 3-isopropylmalate, via the formation of 2-isopropylmaleate.  
CC -1- CATALYTIC ACTIVITY: 3-isopropylmaleate = 2-isopropylmaleate +  
CC H(2)O.  
CC -1- CATALYTIC ACTIVITY: 2-isopropylmaleate + H(2)O = 2-  
CC isopropylmaleate.  
CC -1- PATHWAY: Leucine biosynthesis; second step.  
CC -1- SUBUNIT: Heterodimer of leuc and leud (By similarity).  
CC -1- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY. LEUC 1  
CC SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AP003003; BAB50968.1; -.  
DR InterPro; IPR001030; Aconitase\_N.  
DR InterPro; IPR004430; Leuc.  
DR Pfam; PF00330; aconitase; 1.  
DR PRINTS; PR00415; ACONITASE.  
DR PRODOM; PD000511; Aconitase\_N; 1.  
DR TIGRFAMs; TIGR00170; leuc; 1.  
DR PROSITE; PS00450; ACONITASE\_1; 1.  
DR PROSITE; PS01244; ACONITASE\_2; 1.  
KW Leucine biosynthesis; Lyase; Iron-sulfur; 4Fe-4S; Complete proteome.  
FT METAL 350 350 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
FT METAL 410 410 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
FT METAL 413 413 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
SQ SEQUENCE 469 AA; 50871 MW; 641532D3D4F06888 CRC64;

Query Match 6.3%; Score 83.5; DB 1; Length 469;  
Best Local Similarity 24.8%; Pred. No. 8.9;  
Matches 38; Conservative 27; Mismatches 61; Indels 27; Gaps 6;  
QY 95 GLDIKYNPREVGADRIVNAVAGIHLGSPLIIVDFGTATTTCYINEHKQVMGGAIPGI 154  
Db 97 GVEYSENDIROGIVHIIGPEQFTLPGMTIVCGDSHTST- - - - -HGAF--GALAHGI 147  
QY 155 - - - - -MISTEALFARA- - -LPRIETARPDDIIGKNTVSAM- - - - -QAGILYGYV 197  
Db 148 GTSEVEHVLATQTLIORKAKNMLVRVDGQLPEGVTAKDIILAIIGEIGTAGGTGYVIEYA 207  
QY 198 GQVEGIVSRMKAKSKIPPKVIVATGGLAPLIASE 230  
Db 208 GEAIRSLS-MEGRMTICNMSIEGGARAGLIAAD 239

RESULT 12  
HHP1\_SCHPO  
ID HHP1\_SCHPO STANDARD; PRT; 365 AA.  
AC P40235;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Casein kinase I homolog hhp1 (EC 2.7.1.-).  
GN HHP1 OR SPBC3H7.15.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;





KW Transport; Plasmid; Transmembrane; Chromate resistance.  
FT TRANSMEM 26 46 POTENTIAL.  
FT TRANSMEM 67 87 POTENTIAL.  
FT TRANSMEM 93 113 POTENTIAL.  
FT TRANSMEM 124 144 POTENTIAL.  
FT TRANSMEM 172 192 POTENTIAL.  
FT TRANSMEM 215 235 POTENTIAL.  
FT TRANSMEM 237 257 POTENTIAL.  
FT TRANSMEM 282 302 POTENTIAL.  
FT TRANSMEM 330 350 POTENTIAL.  
FT TRANSMEM 356 376 POTENTIAL.  
FT TRANSMEM 379 399 POTENTIAL.  
SQ SEQUENCE 401 AA; 42642 MW; 6DC963FBADA6DBC5 CRC64;  
  
Query Match 6.3%; Score 83; DB 1; Length 401;  
Best Local Similarity 23.8%; Pred. No. 8.2;  
Matches 46; Conservative 36; Mismatches 75; Indels 36; Gaps 10;  
  
QY 81 FHIKPLIVGPIKGTGLDIKYDNPREVGDRIYNAVAGIHLYGSPLIIVDF---GTATTYC 137  
b 207 FAATPLPAASGMMSTLD--WPLLSQIG---VFFAKAGAFVFGSLAIVPFLYGGVVTEYH 261  
  
QY 138 YINEHKQYMGGA----IAPG-IMISTEALFARAARKLPRIEIAIRPDDIIGKNTVSAMQA-- 190  
Db 262 WLND-KQFVDVAVAVAMITPGPVITVGFVIGLVAGLPACVAAAATFLPCYLFVLPAFY 320  
  
QY 191 ----GILYGVQVEGIVSRMKAISKIPPKVIATGGL--APLIASESDIIDVDPFLTLT 244  
Db 321 FKKYKLPAILAFVDGVTA-----AAIGAITGAVIVLAKRSIVDIPTALLALV 368  
  
QY 245 GLKLL--YEKNT 255  
Db 369 TVALLLKFKLSE 381

RESULT 14  
CIMA\_METJA STANDARD; PRT; 491 AA.  
ID CIMA\_METJA  
AC Q58787;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE (R)-citramalate synthase (EC 4.1.3.-).  
GN CIMA OR MJ1392.  
OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
OC Methanocaldococcaceae; Methanocaldococcus.  
X NCBI\_TaxID=2190;  
N [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;  
RX MEDLINE=96337999; PubMed=8688087;  
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,  
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
jannaschii.";  
RL Science 273:1058-1073(1996).  
RN [2]  
RP SEQUENCE OF 1-10, AND CHARACTERIZATION.  
RX PubMed=9864346;  
RA Howell D.M., Xu H., White R.H.;  
RT "(R)-citramalate synthase in methanogenic archaea.";  
RL J. Bacteriol. 181:331-333(1999).  
CC -!- FUNCTION: Catalyzes the condensation of pyruvate and acetyl-  
coenzyme A to form (R)-citramalate.  
CC -!- SUBUNIT: Homodimer.  
CC -!- SIMILARITY: BELONGS TO THE ALPHA-IPM SYNTHETASE / HOMOCITRATE

CC SYNTHASE FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; U67579; AAB99402.1; -.  
CC TIGR; MJ1392; -.  
CC InterPro; IPR002034; AIPM/Hcit\_synth.  
CC InterPro; IPR000891; HMGL-like.  
CC Pfam; PF00682; HMGL-like; 1.  
CC PROSITE; PS00815; AIPM\_HOMOCIT\_SYNTH\_1; 1.  
CC PROSITE; PS00816; AIPM\_HOMOCIT\_SYNTH\_2; 1.  
KW Lyase; Complete proteome.  
FT CONFLICT 2 2 M -> L (IN REF. 2).  
SQ SEQUENCE 491 AA; 53492 MW; 9D4FA0EA6F8F3BE5 CRC64;

Query Match 6.3%; Score 83; DB 1; Length 491;  
Best Local Similarity 22.4%; Pred. No. 10;  
Matches 57; Conservative 30; Mismatches 71; Indels 96; Gaps 13;  
  
QY 34 KTEDEYGMIMKALLNHVGLQFSDIRGIIISVVPPIMFALERMCLKYFIKPLIVGPGIK 93  
Db 107 KTEDE---VLETALKAV--EYAKEHGLIVELSAEDATRSVDFLIKLF----- 149  
  
QY 94 TGLDIKYNPREVGADRI-----VNAVAGIHLYGSPLIIVD 129  
Db 150 -----NEGEKVGADRVCCVTGVLTTPKQSQELFKKITTENVLPVSVHCHN-----D 196  
  
QY 130 FGTATTYCYINEHKQYMGGAIAPIGIMISTEALFARAARKLPRIEIAIRPDDIIGKNTVSAMQ 189  
Db 197 FGMATA---NTCSAVLGGAVQCHVTNGIGERAGNASLEEV-----VAALK 239  
  
QY 190 AGILYGY-----VGQVEGIVSRMKAISKIPPKVIATGGLAPLIASESDIIDVDPFL 241  
Db 240 --ILYGYDTKIKMEKLYEVSRIVSRL-MKLPVPPNKAIVGDNA--FAHEAGI--HVDGLI 292  
  
QY 242 TLTGLKLLYEKNT 255  
Db 293 -----KNT 296

RESULT 15  
YD43\_METJA STANDARD; PRT; 420 AA.  
ID YD43\_METJA  
AC Q58739;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Putative ammonium transporter MJ1343.  
GN MJ1343.  
OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
OC Methanocaldococcaceae; Methanocaldococcus.  
OX NCBI\_TaxID=2190;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;  
RX MEDLINE=96337999; PubMed=8688087;  
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,  
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
jannaschii.";



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:59:48 ; Search time 22.6949 Seconds  
(without alignments)  
2342.388 Million cell updates/sec

Title: US-09-813-453A-49  
Perfect score: 1321  
Sequence: 1 MIFVLDVGNNTVLGVYDGD.....PFLTTLGKLLYEKNTKKG 258

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL\_21.\*
- 1: sp\_archaea.\*
  - 2: sp\_bacteria.\*
  - 3: sp\_fungi.\*
  - 4: sp\_human.\*
  - 5: sp\_invertebrate.\*
  - 6: sp\_mammal.\*
  - 7: sp\_mhc.\*
  - 8: sp\_organelle.\*
  - 9: sp\_phage.\*
  - 10: sp\_plant.\*
  - 11: sp\_rodent.\*
  - 12: sp\_virus.\*
  - 13: sp\_vertebrate.\*
  - 14: sp\_unclassified.\*
  - 15: sp\_rvirus.\*
  - 16: sp\_bacteriap.\*
  - 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1259	95.3	258	2	Q9F985
2	975	73.8	254	16	Q9KGH5
3	855.5	64.8	259	16	Q8YAC5
4	848.5	64.2	259	16	Q92F54
5	785	59.4	255	16	Q8R7M2
6	767	58.1	259	16	Q8XHL5
7	755	57.2	273	16	Q97EB4
8	663	50.2	265	16	Q9X8N6
9	575	43.5	261	16	Q9A6Z1
10	504.5	38.2	274	16	Q9CD56
11	479.5	36.3	272	16	O06282
12	415.5	31.5	262	16	Q9RX54
13	409	31.0	256	16	Q8RFE4
14	375.5	28.4	246	16	Q9WZY5
15	352.5	26.7	212	2	O32514
16	342.5	25.9	273	16	O83446

17	256.5	19.4	262	16	O51477
18	162	12.3	56	2	P94305
19	158	12.0	295	16	Q8Y2M4
20	150	11.4	257	16	P74045
21	148.5	11.2	229	16	O67753
22	148	11.2	276	16	Q8YQD7
23	146	11.1	592	16	Q9JXF1
24	146	11.1	592	16	Q9JWI7
25	123.5	9.3	248	16	Q9HWC1
26	115	8.7	242	16	Q9PC14
27	105.5	8.0	597	17	O30225
28	105	7.9	224	16	Q98Q93
29	103.5	7.8	512	16	Q8UBH3
30	99.5	7.5	223	16	Q9ZKY6
31	97	7.3	434	16	Q97LQ4
32	96	7.3	394	16	Q8Z0X3
33	95	7.2	5825	10	O82731
34	94	7.1	148	16	Q98627
35	94	7.1	153	16	Q92VL2
36	94	7.1	232	16	Q92P19
37	94	7.1	520	16	Q8U6P6
38	94	7.1	816	16	Q8RB00
39	93.5	7.1	465	5	Q9U2M4
40	93	7.0	422	16	Q8ZA41
41	92.5	7.0	212	16	Q98D86
42	92	7.0	601	16	Q9CPB2
43	91	6.9	394	16	Q8ZJZ6
44	90	6.8	318	10	Q9LVY3
45	90	6.8	405	16	Q98AS0

ALIGNMENTS

RESULT 1  
Q9F985 PRELIMINARY; PRT; 258 AA.  
AC Q9F985;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Putative 32 kDa replication protein.  
OS Bacillus stearothermophilus.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Geobacillus.  
OX NCBI\_TaxID=1422;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=V;  
RA Vasquez C., Pichuanes S., Saavedra C.;  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF198621; AAG28531.1; -  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMs; TIGR00671; baf; 1.  
SQ SEQUENCE 258 AA; 28101 MW; 507B55D695095855 CRC64;

Query Match	95.3%	Score 1259;	DB 2;	Length 258;
Best Local Similarity	94.6%	Pred. No. 6.8e-107;		
Matches 243;	Conservative 8;	Mismatches 6;	Indels 0;	Gaps 0;
Qy	1	MIFVLDVGNNTVLGVYDGDDELKHHWRIETSRSKTEDEYGMKALLNHVGLQFSDIRGI	60	
Db	1	MIFVLDVGNNTVLGVYDGDDELKHHWRIETSRGKTEDEYGMTIKALLNHVGLQFSDIDGI	60	
Qy	61	IISSVVPPIMFALERMCLKYFHFKPLIVGPGIKTGLDIKYDNPREVGADRIVNAVAGIHL	120	
Db	61	IISSVVPPIMFALERMCLKYFHFKPLIVGPGIKTGLNIKYDNPREVGADRIVNAVAGIHL	120	
Qy	121	YGSPLIIVDFGTATTTCYINEHKQYMGGAIAPIGIMISTEALFARAAKLPRIEIRPDDII	180	
Db	121	YGSPLIIVDFGTATTTCYINEHKQYMGGAIAPIGIMISTEALFARAAKLPRIEIRPDDII	180	

QY 181 GKNVTSAMQAGILYGYVGQVEGIVSRMKAISKIPPKVIATGGLAPLIASESDIIDVVDPF 240  
Db 181 GKNVTSAMQAGILYGYVGQVEGIVSRMKAISKIPPKVIATGGLASLIASESNVIDIVDPF 240  
QY 241 LTLTGLKLLYEKNTTEKK 257  
Db 241 LTLTGLKILYEKNVDKK 257

RESULT 2

Q9KGH5  
ID Q9KGH5 PRELIMINARY; PRT; 254 AA.  
AC Q9KGH5;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical protein BH0086.  
GN BH0086.

OS Bacillus halodurans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
C Bacillaceae; Bacillus.  
X NCBI\_TaxID=86665;  
AN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=C-125 / JCM 9153;  
RX MEDLINE=20512582; PubMed=11058132;  
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,  
Horikoshi K.;

RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
halodurans and genomic sequence comparison with Bacillus subtilis.";  
RL Nucleic Acids Res. 28:4317-4331(2000).  
DR EMBL; AP001507; BAB03805.1; -.  
DR InterPro; IPR004619; Baf..  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMS; TIGR00671; baf; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 254 AA; 27907 MW; 0E3B557BA7DAC176 CRC64;

Query Match 73.8%; Score 975; DB 16; Length 254;  
Best Local Similarity 70.8%; Pred. No. 5.8e-81;  
Matches 179; Conservative 36; Mismatches 38; Indels 0; Gaps 0;

QY 1 MIFVLDVGNNTNTVLGVYDGDDELKHHWRIETSRSKTEDEYGMKALLNHVGLQFSDIRGI 60  
Db 1 MILVIDVGNNTNTVLGVYQDETLLVHHWRLATSRQKTEDEYAMTVRSLEFDHAGLQFQDIDGI 60  
QY 61 IISVVPPIMFALERMCLKYFHFKPLIVGPGIKTGLDIKYNPREVGADRIVNAVAGIHL 120  
61 VISSVPPMFSLEQMCKKYFHTVPMIIGPGIKTGLNLIKYNPKVEGADRIVNAVAIEL 120  
QY 121 YGSPLIIVDFGTATTTCYINEHKQYMGGAIPGIMISTEALFARAAKLPRIEIAFPDDII 180  
Db 121 YGYPPIVVDFTATTTCYINEHKQYMGGAIPGIMISTEALFARAAKLPRIEIAFPDDII 180  
QY 181 GKNVTSAMQAGILYGYVGQVEGIVSRMKAISKIPPKVIATGGLAPLIASESDIIDVVDPF 240  
Db 181 GTNTIDSMQSGIFYGYVSQDGVVGRMKAQAESEPKVIATGGLAKLIGTSETIDVIDSF 240

QY 241 LTLTGLKLLYEKN 253  
Db 241 LTLKGLQLIYKKN 253

RESULT 3

Q8YAC5  
ID Q8YAC5 PRELIMINARY; PRT; 259 AA.  
AC Q8YAC5;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical protein lmo0221.  
GN LMO0221.

OS Listeria monocytogenes.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Listeriaceae; Listeria.  
OX NCBI\_TaxID=1639;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=EGD-E / SEROVAR 1/2A;  
RX MEDLINE=21537279; PubMed=11679669;  
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,  
Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,  
Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,  
Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,  
Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
Rommel B., Rose M., Schlueter T., Simoes N., Tierrez A.,  
Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;  
RT "Comparative genomics of Listeria species.";  
RL Science 294:849-852(2001).  
DR EMBL; AL591974; CAD00748.1; -.  
DR MEROPS; M41.009; -.  
DR ListiList; LMO00221; -.  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMS; TIGR00671; baf; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 259 AA; 28187 MW; F106B049D80025B7 CRC64;

Query Match 64.8%; Score 855.5; DB 16; Length 259;  
Best Local Similarity 62.9%; Pred. No. 4.9e-70;  
Matches 163; Conservative 41; Mismatches 52; Indels 3; Gaps 1;

QY 1 MIFVLDVGNNTNTVLGVYDGDDELKHHWRIETSRSKTEDEYGMKALLNHVGLQFSDIRGI 60  
Db 1 MILVIDVGNNTCTGVYEQKLLKHHWRTDRHRTSDELGMTVLNFFSYANLTPSDIQGI 60  
QY 61 IISVVPPIMFALERMCLKYFHFKPLIVGPGIKTGLDIKYNPREVGADRIVNAVAGIHL 120  
Db 61 IISVVPPIMHAMETMCVRYFNIRPLIVGPGIKTGLNKLVDNPREIGSDRIVNAVAASEE 120  
QY 121 YGSPLIIVDFGTATTTCYINEHKQYMGGAIPGIMISTEALFARAAKLPRIETARPDDII 180  
Db 121 YGTPVIVVDFTATTTCYIDEISGVYQGGALAPGIMISTEALYNRAAKLPKPRVDIAESSQII 180  
QY 181 GKNVTSAMQAGILYGYVGQVEGIVSRMKAISKIPPKVIATGGLAPLIASESDIIDVVDPF 240  
Db 181 GKSTVSSMQAGIFYGYVQCEGIIAEMKKQSNASPVVWATGGLARMITERKSSAVDILDPF. 240  
QY 241 LTLTGLKLLYEKN---TEK 256  
Db 241 LTLKGLLELYRRNKPTTEK 259

RESULT 4

Q92F54  
ID Q92F54 PRELIMINARY; PRT; 259 AA.  
AC Q92F54;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical protein lin0253.  
GN LIN0253.  
OS Listeria innocua.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Listeriaceae; Listeria.  
OX NCBI\_TaxID=1642;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CLIP 11262 / SEROVAR 6A;  
RX PubMed=11679669;  
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,







Q9A6Z1 PRELIMINARY; PRT; 261 AA.  
AC Q9A6Z1;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Transcriptional activator, putative, Baf family.  
GN CC1935.  
OS Caulobacter crescentus.  
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
OC Caulobacter.  
OX NCBI\_TaxID=155892;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 19089 / CB15;  
RX MEDLINE=21173698; PubMed=11259647;  
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,  
Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,  
Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,  
Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
RT "Complete genome sequence of Caulobacter crescentus.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
DR EMBL; AE005867; AAK23910.1; -;  
DR TIGR; CC1935; -;  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMs; TIGR00671; baf; 1.  
KW Complete proteome.  
SQ SEQUENCE 261 AA; 27965 MW; C19E60D7B0714EF5 CRC64;  
  
Query Match 43.5%; Score 575; DB 16; Length 261;  
Best Local Similarity 44.1%; Pred. No. 2.1e-44;  
Matches 113; Conservative 52; Mismatches 89; Indels 2; Gaps 1;  
  
QY 1 MIFVLDVGNNTVVGVDGDELKHHWRIETSRKTEDEYGMKALNHNHVGQLQFSDIRGI 60  
Db 2 MLLAIEQNTNTMFAIHGASWVAQWRSATESTRTADEYVWLSQLLSMQGLGFRDAIDAV 61  
  
QY 61 IISVVPPIMFALERMCLKYFHKPLIVPGIKTGLDIKYNPREVGADRIVNAVAGIHL 120  
Db 62 IISVVPPQSIENLNLRRYFNVEPLVIGENAKLIGDRIEKPSEAGADRLVNAIGAMV 121  
  
QY 121 YGSLPIIVDFGTATTTCYINEHKQYMGGAIAPIGIMISTEALFARAAPRIEIRP--DD 178  
Db 122 YPGPLVIDSGTATTFDIVAADGAFEGGIAPGINLSMQALHEAAKLPRIAIQRPAGNR 181  
  
QY 179 IIGKNTVSAMQAGILYGVQVEGIVSRMKAKSKIPPVATGGLAPLIASESDIIDVVD 238  
Db 182 IVGTDTVSAMQSGVFWGYISLIEGLVARIKAERGEPTVIATGGVASLFEGATDSIDHFD 241  
  
QY 239 PFLTTLGLKLLYEKNT 254  
Db 242 SDLTIRGLLEIYRRNT 257  
  
RESULT 10  
Q9CD56 PRELIMINARY; PRT; 274 AA.  
ID Q9CD56  
AC Q9CD56;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical protein ML0232.  
GN ML0232.  
OS Mycobacterium leprae.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1769;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=TN;  
RX MEDLINE=21128732; PubMed=11234002;  
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,  
Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser J., Hamlin N.,  
Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,  
Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
Barrell B.G.;  
RT "Massive gene decay in the leprosy bacillus.";  
RL Nature 409:1007-1011(2001).  
DR EMBL; AL583917; CAC29740.1; -;  
DR Leproma; ML0232; -;  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMs; TIGR00671; baf; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 274 AA; 29421 MW; 1C2E735BDEC78765 CRC64;  
  
Query Match 38.2%; Score 504.5; DB 16; Length 274;  
Best Local Similarity 38.7%; Pred. No. 6.1e-38;  
Matches 105; Conservative 58; Mismatches 85; Indels 23; Gaps 6;  
  
QY 1 MIFVLDVGNNTVVGVDGDELKHH-----WRIETSRKTEDEYGMKALNHNHVGQLQF 54  
Db 1 MLLAIDVRNTHTVVGLLSGS--KEHAKVVQWRIETSEVTADELALIIDGL---IGDDS 55  
  
QY 55 SDIRGIISVVPPIMFALERMCLKYFHKPLIVPGIKTGLDIKYNPREVGADRIVN 113  
Db 56 ERLAGAAALSTVPSVLHEVRIMLDQYWPSPVHVLIEPGVRTGIPLLVDNPKVGDRIVN 115  
  
QY 114 AVAGIHLGSLPIIVDFGTATTTCYINEHKQYMGGAIAPIGIMISTEALFARAAPRIE 173  
Db 116 CLAAFKFGQAAIIVDFGSSICVDVWSAKGEFLGGAIAPIGVSSDAAAARSALRRVEL 175  
  
QY 174 ARPDDIIGKNTVSAMQAGILYGVQVEGIVSRMKAK-----SKIPPVATGGLA 224  
Db 176 APRSRVVGKNTVECMQAGVFGFAGLVGDLVGRMRQDVEEFSGDLGNRV--AVVATGHTA 233  
  
QY 225 PLIASESDIIDVVDPFLLTLGLKLLYEKNT 255  
Db 234 PLLLPHELTVVDHYDRHLTLHGLRLVFERNRE 264  
  
RESULT 11  
O06282 PRELIMINARY; PRT; 272 AA.  
ID O06282  
AC O06282;  
DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical 29.3 kDa protein (Transcriptional activator, putative,  
DE Baf family).  
GN RV3600C OR MTCY07H7B.22 OR MT3706.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the



```
RT complete genome sequence." ;
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / OSHKOSH;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains." ;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z95557; CAB08944.1; -.
DR EMBL; AE007170; AAK48063.1; -.
DR TIGR; MT3706; -.
DR TubercuList; RV3600c; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00671; baf; 1.
W Hypothetical protein; Complete proteome.
Q SEQUENCE 272 AA; 29304 MW; 5D70E6FEF0F09AC8B CRC64;

Query Match 36.3%; Score 479.5; DB 16; Length 272;
Best Local Similarity 37.8%; Pred. No. 1.2e-35;
Matches 101; Conservative 56; Mismatches 93; Indels 17; Gaps 5;

QY 1 MIFVLDVGNNTVVLGVYDGDDELKHH-----WRIETSRKTEDEYGMKIKALLNHVGLQF 54
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MLLAIDVRNTHVTVGLLSG--MKEHAKVQVQWRIRTESEVTADELALTIDGL---IGEDS 55

QY 55 SDIRGIISVVPPIMFALERMLKYPHIKP-LIVPGIKTGLDIKYDNPKEVGADRIVN 113
   : | | | : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 56 ERLTGAALSTVPSVLHEVRIMLDQVWPSPVPHVLIETPGVRTGIPLLVDNPKVEGADRIVN 115

QY 114 AVAGIHLYGSLIIVDFGTATTTCYINEHKQYMGGAIAPIGIMISTEALFARAACLPRIEI 173
   : | : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 116 CLAAIDRFRKAAIIVDFGSSICVDVVSAGKEFLGGAIAPIGVQVSSDAAAARSALRRVEL 175

QY 174 ARPDDIIGKNTVSAMQAGILYGVGVQVEGIVSRMKA-----KSKIPPKVIATGGLAPLIA 228
   ||| : : : : : ||| : : : : : : : : : : : : : : : : : : : :
Db 176 ARPRSIVGKNTVECMQAGAVFGFAGLVGLVGRIREDSVGSFVDHVAIVATGHTAPLLL 235

QY 229 SESDIIDVVDPEFLTTLTGLKLLYEKNTE 255
   | : | | | | | | | | | | : : : : :
Db 236 PELHTVDHYDQHLTLQGLRLVFERNLE 262

RESULT 12
3RX54
AC Q9RX54 PRELIMINARY; PRT; 262 AA.
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical protein DR0461.
GN DR0461.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
```

```
RT radiodurans R1." ;
RL Science 286:1571-1577(1999).
DR EMBL; AE001905; AAF10040.1; -.
DR TIGR; DR0461; -.
DR InterPro; IPR004619; Baf.
DR InterPro; IPR001220; Lectin_legB.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00671; baf; 1.
DR PROSITE; PS00307; LECTIN_LEGUME_BETA; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 262 AA; 27839 MW; 965EAD2F78785A0 CRC64;

Query Match 31.5%; Score 415.5; DB 16; Length 262;
Best Local Similarity 35.2%; Pred. No. 7.8e-30;
Matches 94; Conservative 49; Mismatches 99; Indels 25; Gaps 6;

QY 2 IFVLDVGNNTVVLGVYDGD-ELKHHWRIETSRKTEDEYGMKIKALLNHVGLQFSDIRGI 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 6 LLAVDIGNTTTLGLADASGALTHTWIRTNREMLPDDLALQLHGLFTLAGAPIP--RAA 63

QY 61 IISVVVPPI-----MFALERMLKYPHIKPLIVPGIKTGLDIKYDNPKEVGADRIVNAVA 116
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 64 VLSSVAPPVGENYALAKR---HFMDAFAVSAENLPDVTVELDTPGSVGADRLCN--- 116

QY 117 GIHLYGSP-----LIIVDFGTATTTCYINEHKQYMGGAIAPIGIMISTEALFARAACL 168
   | : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 117 ---LFGAEKYLGLDYAVVDFGTSTNFDVVGRGRFLGILATGAQVSADALFARAACL 173

QY 169 PRIEIARPDDIIGKNTVSAMQAGILYGVGVQVEGIVSRMKAASKIPPKVIATGGLAPLIA 228
   ||| : : : : : ||| : : : : : : : : : : : : : : : : : : : :
Db 174 PRITLOAPETAIGKNTVHALQSLVFGYAEVMDGLLRIRRAELPGEAVAVATGGFSRTVQ 233

QY 229 SESDIIDVVDPEFLTTLTGLKLLYEKNTE 255
   || | | | | | | | | | : :
Db 234 GICQEIDYDETLTLRGLVELWASRSE 260

RESULT 13
Q8RFE4
ID Q8RFE4 PRELIMINARY; PRT; 256 AA.
AC Q8RFE4;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Bvg accessory factor.
GN FN0761.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasieva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Fonstein M., Kyrpides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586." ;
RL J. Bacteriol. 184:2005-2018(2002).
DR EMBL; AE010586; AAL94957.1; -.
KW Complete proteome.
SQ SEQUENCE 256 AA; 28601 MW; CDF681127F08914B CRC64;

Query Match 31.0%; Score 409; DB 16; Length 256;
Best Local Similarity 37.5%; Pred. No. 3e-29;
Matches 96; Conservative 46; Mismatches 110; Indels 4; Gaps 4;

QY 1 MIFVLDVGNNTVVLGVYDGD-ELKHHWRIETSRKTEDEYGMKIKALLNHVGLQFSDIRG 59
   || : : : : : | : : : : : : : : : : : : : : : : : : : :
Db 1 MIIGIDIGNTHIVTGIYDNGGELISTFRIATNDKMTDEYFSYFNITKYNEISIKKVDA 60
```



QY 60 IILSSVPPIMFALERMCLKYFHIKPLIVGPGIKTGLDI-KYDNPREVGAADRVNAVAGI 118  
:|||||: : |||: : |||: : |||: : |||: :  
Db 61 IILSSVPPNIIITQFFARKYFKVEATIVDLEKKLPFTFAKGINYTGFGADRIIDTEAM 120  
:|||||: : |||: : |||: : |||: : |||: :  
QY 119 HLY-GSPLIIVDFGTATTCYINEHKQYMGGAIAPIAGIMISTEALFARAACKLPRIETARP 177  
:|||||: : |||: : |||: : |||: : |||: :  
Db 121 QKYPDKNLVIFDFGTATYDVLKK-GVYIGGGILPGIDMSINALYNGTAKLPVRKFTTPS 179  
:|||||: : |||: : |||: : |||: : |||: :  
QY 178 DIIGKNTVSAMQAGILYGVGVQVEGIVSRMKAKSKIPPVKVIATGGLAPLIASESDIIDV 237  
:|||||: : |||: : |||: : |||: : |||: :  
Db 180 SVLGTDTMKQIQAAIFGYAGQIKHIIKKINEELNEEFVLATGGLGKILSAEIDEIDEY 239  
:|||||: : |||: : |||: : |||: : |||: :  
QY 238 DPFLTLTGLKLLYEKN 253  
:|||||: : |||: : |||: : |||: : |||: :  
Db 240 DANLSLKGlyTLyKLN 255  
:|||||: : |||: : |||: : |||: : |||: :

RESULT 14  
Q9WZY5 PRELIMINARY; PRT; 246 AA.  
Q9WZY5;  
DT 01-NOV-1999 (TReMBLrel. 12, Created)  
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Hypothetical protein TM0883.  
GN TM0883.  
OS Thermotoga maritima.  
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.  
OX NCBI\_TaxID=2336;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MSB8 / DSM 3109;  
RX MEDLINE=99287316; PubMed=10360571;  
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,  
Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
McDonald L., Otterback T.R., Malek J.A., Linher K.D., Garrett M.M.,  
Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
Saizberg S.L., Smith H.O., Venter J.C., Fraser C.M.;  
RT "Evidence for lateral gene transfer between Archaea and Bacteria from  
genome sequence of Thermotoga maritima."  
RL Nature 399:323-329(1999).  
DR EMBL; AE001754; 27154 MW; 9E0309AD462CF266 CRC64;  
DR TIGR; TM0883;  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMs; TIGR00671; baf; 1.  
W Hypothetical protein; Complete proteome.  
Q SEQUENCE 246 AA; 27154 MW; 9E0309AD462CF266 CRC64;

Query Match 28.4%; Score 375.5; DB 16; Length 246;  
Best Local Similarity 37.5%; Pred. No. 3.2e-26;  
Matches 94; Conservative 48; Mismatches 92; Indels 17; Gaps 9;  
QY 1 MIFVLDVGNNTVLGVYDDELKHHWRITRSKTEDEYGMKALLNHVGLQFSDIRGI 60  
:|||||: : |||: : |||: : |||: : |||: :  
Db 1 MYLLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDELFSLHPLL---GDAMREIKGI 57  
:|||||: : |||: : |||: : |||: : |||: :  
QY 61 IISVVPPIMFALERMCLKYFHIKPLIVGPGIKTGLDIKYD--NPREVGAADRVNAVAGI 118  
:|||||: : |||: : |||: : |||: : |||: :  
Db 58 GVASVPTQNTVIERFSQKYFHISPIW--KAKNGC-VKNVKNPSEVGADRVANVAVFV 114  
:|||||: : |||: : |||: : |||: : |||: :  
QY 119 HLYGSPLIIVDFGTATTCYINEHKQYMGGAIAPIAGIMISTEALFARAACKLPRIETARP 178  
:|||||: : |||: : |||: : |||: : |||: :  
Db 115 KEYGKNGI IIDMGTATVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKLPVEVKPADF 173  
:|||||: : |||: : |||: : |||: : |||: :  
QY 179 IIGKNTVSAMQAGILYGVGVQVEGIVSRMK-AKSKIPPVKVIATGGLAPLIASESDII--D 235  
:|||||: : |||: : |||: : |||: : |||: :  
Db 174 VVGKDTENIRLGVVNGSVYALEGIIGRIKEIVGDLP--VVLTTGGQSKIV---KDMIKHE 228  
:|||||: : |||: : |||: : |||: : |||: :  
QY 236 VVDPFLTLTGL 246  
:|||||: : |||: : |||: : |||: : |||: :  
Db 229 IFDEDLTIKGV 239  
:|||||: : |||: : |||: : |||: : |||: :

RESULT 15  
O32514 PRELIMINARY; PRT; 212 AA.  
ID O32514  
AC O32514;  
DT 01-JAN-1998 (TReMBLrel. 05, Created)  
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Similar to Bacillus subtilis.  
OS Desulfovibrio vulgaris (strain Miyazaki).  
OC Bacteria; Proteobacteria; delta subdivision; Desulfovibrionaceae;  
OC Desulfovibrio.  
OX NCBI\_TaxID=883;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MIYAZAKI;  
RA Kitamura M., Konishi T., Kawanishi K., Ohashi K., Kishida Y.,  
Kohno K., Akutsu H., Kumagai I., Nakaya T.;  
RT "Sequence analyses of two ferredoxin genes and their flanking regions  
from Desulfovibrio vulgaris (Miyazaki F).";  
RL J. Biochem. Mol. Biol. Biophys. 2:147-154(1998).  
DR EMBL; AB005550; BAA21476.1;  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
SQ SEQUENCE 212 AA; 22336 MW; 23E1789973A344D7 CRC64;  
Query Match 26.7%; Score 352.5; DB 2; Length 212;  
Best Local Similarity 37.7%; Pred. No. 3.3e-24;  
Matches 78; Conservative 42; Mismatches 80; Indels 7; Gaps 5;  
QY 4 VLDVGNNTVLGVYDDELKHHWRITRSKTEDEYGMKALLNHVGLQFSDIRGIIS 63  
:|||||: : |||: : |||: : |||: : |||: :  
Db 7 LFDIGNTNVKIGIAVETAVLTSTVLPDTPGQTTDSIGLRLLLEVLRHAGLGPADVGCACVAS 66  
:|||||: : |||: : |||: : |||: : |||: :  
QY 64 SVVPPIMFALERMCLKYFHIKPLIVGPG-IKTGLDIKYDNPREVGAADRVNAVAGIHLYG 122  
:|||||: : |||: : |||: : |||: : |||: :  
Db 67 SVVPGVNPLIRRACERYLYRK-LLFAPGDIAIPLDNRYPAEVGAADRLVAAYAAARRLYP 125  
:|||||: : |||: : |||: : |||: : |||: :  
QY 123 SP--LIIVDFGTATTCYINEHKQYMGGAIAPIAGIMISTEALFARAACKLPRI--EIARPD 178  
:|||||: : |||: : |||: : |||: : |||: :  
Db 126 GPRSLVSVDFGTATTFDCV-EGGAYLGLGICPGVLSSAGALSSRTAKLPRLISLEVEEDSP 184  
:|||||: : |||: : |||: : |||: : |||: :  
QY 179 IIGKNTVSAMQAGILYGVGVQVEGIVS 205  
:|||||: : |||: : |||: : |||: : |||: :  
Db 185 VIGRSTTSLNHGFIFGFAAMTEGVLA 211  
:|||||: : |||: : |||: : |||: : |||: :

Search completed: June 24, 2003, 22:16:13  
Job time : 23.6949 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run On: June 24, 2003, 21:49:47 ; Search time 10.4678 Seconds  
(without alignments)  
725.188 Million cell updates/sec

Title: US-09-813-453A-49  
Perfect score: 1321  
Sequence: 1 MIFVLDVGNNTVLGVYDGD.....PFLTLTGLKLLYEKNTTEKKG 258

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92	7.0	330	4	US-09-134-001C-3582
2	84	6.4	342	4	US-08-096-181A-12
3	84	6.4	342	4	US-08-096-181A-14
4	84	6.4	342	5	PCT-US94-08326-12
5	84	6.4	342	5	PCT-US94-08326-14
6	84	6.4	361	4	US-08-096-181A-8
7	84	6.4	361	5	PCT-US94-08326-8
8	84	6.4	363	4	US-08-096-181A-10
9	84	6.4	363	5	PCT-US94-08326-10
10	83	6.3	365	1	US-08-447-500-4
11	83	6.3	365	1	US-08-454-097-4
12	83	6.3	365	1	US-08-453-866-4
13	83	6.3	365	3	US-09-100-664A-13
14	83	6.3	365	3	US-08-185-359-4
15	83	6.3	365	4	US-09-335-983-13
16	82	6.2	314	4	US-09-134-001C-2983
17	81.5	6.2	494	1	US-08-447-500-2
18	81.5	6.2	494	1	US-08-454-097-2
19	81.5	6.2	494	1	US-08-447-408-2
20	81.5	6.2	494	1	US-08-453-866-2
21	81.5	6.2	494	3	US-08-185-359-2
22	81.5	6.2	494	4	US-09-457-040B-25
23	79.5	6.0	421	4	US-09-134-001C-5201
24	79	6.0	489	4	US-09-431-099-4
25	79	6.0	489	4	US-09-431-099-4
26	78.5	5.9	1012	2	US-08-475-891A-4
27	78.5	5.9	1025	2	US-08-567-375-4

28	78.5	5.9	1025	2	US-08-587-680A-4	Sequence 4, Appli
29	76.5	5.8	201	2	US-08-911-364-2	Sequence 2, Appli
30	76.5	5.8	731	2	US-08-911-364-1	Sequence 1, Appli
31	76.5	5.8	733	4	US-08-464-700-2	Sequence 2, Appli
32	76	5.8	248	4	US-09-134-001C-3211	Sequence 3211, Ap
33	76	5.8	399	4	US-09-134-001C-4571	Sequence 4571, Ap
34	75.5	5.7	542	3	US-08-968-563-16	Sequence 16, Appl
35	75.5	5.7	542	4	US-08-969-683A-16	Sequence 16, Appl
36	75.5	5.7	542	4	US-09-297-928-12	Sequence 12, Appl
37	74.5	5.6	609	4	US-08-969-683A-67	Sequence 67, Appl
38	74.5	5.6	1375	4	US-09-171-410-1	Sequence 1, Appli
39	73.5	5.6	340	4	US-09-789-300A-2	Sequence 2, Appli
40	73.5	5.6	625	1	US-08-391-615-3	Sequence 3, Appli
41	73	5.5	724	3	US-08-793-331-4	Sequence 4, Appli
42	73	5.5	725	3	US-08-793-331-6	Sequence 6, Appli
43	73	5.5	2052	3	US-09-045-201A-2	Sequence 2, Appli
44	73	5.5	2052	4	US-09-619-062-2	Sequence 2, Appli
45	72.5	5.5	521	1	US-08-915-003-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1  
US-09-134-001C-3582  
; Sequence 3582, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 3582  
; LENGTH: 330  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3582

Query Match	7.0%	Score 92;	DB 4;	Length 330;
Best Local Similarity	20.6%	Pred. No. 0.014;		
Matches	60;	Conservative	40;	Mismatches 87; Indels 104; Gaps 16;
QY	1	MIFVLDVGNNTVLGVYDGD-ELKHHWRIETSRKTEDEYGMKALLNHVGLQFSDIRG	59	
Db	6	IILAADIGGTTCKLGIFDKDLQLHKWSIDTD---TSDHTG---ELLKNIYNSFT----	55	
QY	60	IIISVVPPIMFALERMLCKLYFIKPLIVPGIKTGLDIKYDNPREVGAIRIVNAVAGIH	119	
Db	56	-----EKIA-----EYKYDFNNVVG-----VGIG	74	
QY	120	LYGSPLIIVDFGTATTTCYINEH-----KQYMGGAIPGIMISTEALFARAACL	168	
Db	75	VPGP---VDFDTGVVYGAVNLHPDVSVNREIFKQYVN---CP-VYVDNDANVAALGEK	126	
QY	169	PRIEIRPDDIIGKNTVSAMQAGIL-----YGYVGQVEGIVSRMKA-----KS	211	
Db	127	HKGAGEGADDVVAITLGTGLGGGIISNGEIVHGHNGS-GAEIGHLRADFQRFQCNCGKS	185	
QY	212	KIPPKVIATGGLAPLIASESDIIDVVD---PFLTLTG---LKLYEKNTTEKK	257	
Db	186	-----GCIETVASATGVVNLVNFYYPKLTFFKSSILQLIKDNQVTAK	226	

RESULT 2  
US-08-096-181A-12  
; Sequence 12, Application US/08096181A







Db 176 KREGAKGENKRPNDKAGEVRIG---EINNGIQVGAKYDANDIVAKIAYG-----R 222  
QY 134 TTYCY--INEHKQYMGGAIAAP-----GIMISTEALFAR 164  
Db 223 TNYKYNESDEHKQQLNGVLATLGYRFSDLGLVSLDSGYAK 263

RESULT 7  
PCT-US94-08326-8  
; Sequence 8, Application PC/TUS9408326  
; GENERAL INFORMATION:  
; APPLICANT: North American Vaccine, Inc.  
; APPLICANT: 12103 Indian Creek Court  
; APPLICANT: Beltsville, MD 20705  
; APPLICANT: Pullen, Jeffrey K.  
; APPLICANT: Soper, Thomas S.  
; APPLICANT: Liang, Shu-Mei  
; TITLE OF INVENTION: A Method For The High Level  
; TITLE OF INVENTION: Expression,  
; TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane  
; TITLE OF INVENTION: Protein  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
; STREET: 1100 New York Avenue, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/08326  
; FILING DATE: Herewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/096,181  
; FILING DATE: 23-JULY-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Esmond, Robert W.  
; REFERENCE/DOCKET NUMBER: 1438.001PC01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 361 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
PCT-US94-08326-8

Query Match 6.4%; Score 84; DB 5; Length 361;  
Best Local Similarity 23.0%; Pred. No. 0.15;  
Matches 37; Conservative 30; Mismatches 54; Indels 40; Gaps 7;  
QY 28 IETSRKT-----EDEYGM-----IKALLNHVGLQFSDIRGIISVVPPIMFAL 73  
Db 119 VKLGRAKTIADGITSADKEYGVLNNSDIPTSGNTVGYTFKIDGLVLGA---NYLLAQ 175  
QY 74 ERMCLKYFHKPLIVGPGIKTGLDIKYDNPREGADRIVNAVAGIHLYGSPLIIVDFGTA 133  
Db 176 KREGAKGENKRPNDKAGEVRIG---EINNGIQVGAKYDANDIVAKIAYG-----R 222  
QY 134 TTYCY--INEHKQYMGGAIAAP-----GIMISTEALFAR 164  
Db 223 TNYKYNESDEHKQQLNGVLATLGYRFSDLGLVSLDSGYAK 263

RESULT 8  
US-08-096-181A-10  
; Sequence 10, Application US/08096181A  
; Patent No. 6153406  
; GENERAL INFORMATION:  
; APPLICANT: Tai, Joseph Y.  
; APPLICANT: Pullen, Jeffrey K.  
; APPLICANT: Soper, Thomas S.  
; APPLICANT: Liang, Shu-Mei  
; TITLE OF INVENTION: A Method For The High Level Expression,  
; TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane Protein  
; TITLE OF INVENTION: P2 From Haemophilus Influenzae Type b  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
; STREET: 1100 New York Avenue, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/096,181A  
; FILING DATE: 23-Jul-1993  
; CLASSIFICATION: 424  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 363 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-096-181A-10

Query Match 6.4%; Score 84; DB 4; Length 363;  
Best Local Similarity 23.0%; Pred. No. 0.15;  
Matches 37; Conservative 30; Mismatches 54; Indels 40; Gaps 7;  
QY 28 IETSRKT-----EDEYGM-----IKALLNHVGLQFSDIRGIISVVPPIMFAL 73  
Db 121 VKLGRAKTIADGITSADKEYGVLNNSDIPTSGNTVGYTFKIDGLVLGA---NYLLAQ 177  
QY 74 ERMCLKYFHKPLIVGPGIKTGLDIKYDNPREGADRIVNAVAGIHLYGSPLIIVDFGTA 133  
Db 178 KREGAKGENKRPNDKAGEVRIG---EINNGIQVGAKYDANDIVAKIAYG-----R 224  
QY 134 TTYCY--INEHKQYMGGAIAAP-----GIMISTEALFAR 164  
Db 225 TNYKYNESDEHKQQLNGVLATLGYRFSDLGLVSLDSGYAK 265

RESULT 9  
PCT-US94-08326-10  
; Sequence 10, Application PC/TUS9408326  
; GENERAL INFORMATION:  
; APPLICANT: North American Vaccine, Inc.  
; APPLICANT: 12103 Indian Creek Court  
; APPLICANT: Beltsville, MD 20705  
; APPLICANT: Pullen, Jeffrey K.  
; APPLICANT: Soper, Thomas S.  
; APPLICANT: Liang, Shu-Mei  
; TITLE OF INVENTION: A Method For The High Level  
; TITLE OF INVENTION: Expression,  
; TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane  
; TITLE OF INVENTION: Protein  
; TITLE OF INVENTION: P2 From Haemophilus Influenzae Type b



NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1100 New York Avenue, Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/08326  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/096,181  
FILING DATE: 23-JULY-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Esmond, Robert W.  
REFERENCE/DOCKET NUMBER: 1438.001PC01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 363 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US94-08326-10  
Query Match 6.4%; Score 84; DB 5; Length 363;  
Best Local Similarity 23.0%; Pred. No. 0.15;  
Matches 37; Conservative 30; Mismatches 54; Indels 40; Gaps 7;  
QY 28 IETSRSKT-----EDYGM-----IKALLNHVGLQFSDIRGIISSVPPINFAL 73  
Db 121 VKLGRAKTIADGITSADREYGVNLNSDYIPTSGNTVGYTFKGIDGLVGA---NYLLAQ 177  
QY 74 ERMCLKYFHKPLIVPGIKTGLDIKYNPREVGADRIYNAVAGIHLGSPLIIVDFGTA 133  
Db 178 KREGAKGENKRPNDKAGEVRIG---EINNGIQVGAKYDANDIVAKIAYG-----R 224  
QY 134 TTYCY--INEHKQYMGGAIAAP-----GIMISTEALFAR 164  
Db 225 TNYKYNESDEHKQQLNGVLATLGYRFSDLGLLVSLDSGYAK 265  
RESULT 10  
US-08-447-500-4  
Sequence 4, Application US/08447500  
Patent No. 5627064  
GENERAL INFORMATION:  
APPLICANT: Hoekstra, Merl F.  
TITLE OF INVENTION: PROTEIN KINASES  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Spensley Horn Jubas & Lubitz  
STREET: 1880 Century Park East, Suite 500  
CITY: Los Angeles  
STATE: CA  
COUNTRY: USA  
ZIP: 90067  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/447,500

FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/008,001  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Wetherell Ph.D., John R.  
REGISTRATION NUMBER: 31,678  
REFERENCE/DOCKET NUMBER: PD-2458  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 455-5100  
TELEFAX: (619) 455-5110  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 365 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-447-500-4  
Query Match 6.3%; Score 83; DB 1; Length 365;  
Best Local Similarity 23.0%; Pred. No. 0.2;  
Matches 51; Conservative 47; Mismatches 92; Indels 32; Gaps 11;  
QY 16 VYDGBELKHHRIETSRK---TEDEYGMKALLNHVGLQFSDIRGI-----IISVV 66  
Db 31 VVSGEVA--IKLESTRAKHPQLEYEY-RVYRILSGVGIPFVRWFGVECDYNAMVMDLL 87  
QY 67 PPIMFALERMLKLYFHIKP-LIVGPGIKTGLDIKYNP---REVGADRIYNAVAGIHLG 122  
Db 88 GPSLEDLNFENCRKFSKLTVLLADQLISRIEFHSHKSFHLRDIKPD---NFLMGIGKRG 144  
QY 123 SPLIIVDFGTATTTCYINEHKQYMGGAIAAPGIMISTEALFARAALKPRIETARPDDIIGK 182  
Db 145 NQVNIIDFGLAKKY---RDHKTHLHIPYRENKNTLTGTARYASINTHLGIEQSRDDLESL 201  
QY 183 NTVSAMQAGILYGVQV--EGIVSRMKAK--SKIPPKVIAT 220  
Db 202 GYV-----LVYFCRGSPLPWQGLKATTKKQKYEKIMEKKIST 237  
RESULT 11  
US-08-454-097-4  
Sequence 4, Application US/08454097  
Patent No. 5686412  
GENERAL INFORMATION:  
APPLICANT: Hoekstra, Merl F.  
TITLE OF INVENTION: Protein Kinases  
NUMBER OF SEQUENCES: 57  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/454,097  
FILING DATE: 30-MAY-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/185,359  
FILING DATE: 21-JAN-1994  
APPLICATION NUMBER: US 08/008,001  
FILING DATE: 21-JAN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/728,783  
FILING DATE: 03-JUL-1991





**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:24:12 ; Search time 14.3382 Seconds  
(without alignments)  
1947.059 Million cell updates/sec

Title: US-09-813-453A-49  
Perfect score: 1321  
Sequence: 1 MIFVLDVGNNTVLGVYDGD.....PFLTGLKLLYEKNTTEKKG 258

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 417779 seqs, 108206813 residues  
Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1321	100.0	258	9 US-09-813-453A-49	Sequence 49, Appl
2	1057.5	80.1	262	9 US-09-813-453A-45	Sequence 45, Appl
3	1046	79.2	258	9 US-09-813-453A-2	Sequence 2, Appl
4	975	73.8	254	9 US-09-813-453A-47	Sequence 47, Appl
5	900	68.1	233	9 US-09-813-453A-17	Sequence 17, Appl
6	834	63.1	255	9 US-09-813-453A-7	Sequence 7, Appl
7	817.5	61.9	256	9 US-09-813-453A-55	Sequence 55, Appl
8	663	50.2	265	9 US-09-813-453A-4	Sequence 4, Appl
9	629.5	47.7	250	9 US-09-813-453A-3	Sequence 3, Appl
10	575	43.5	260	9 US-09-813-453A-51	Sequence 51, Appl
11	542	41.0	258	9 US-09-813-453A-6	Sequence 6, Appl
12	513	38.8	219	9 US-09-813-453A-57	Sequence 57, Appl
13	479.5	36.3	272	9 US-09-813-453A-5	Sequence 5, Appl
14	475.5	36.0	262	9 US-09-712-363-276	Sequence 276, App
15	415.5	31.5	272	9 US-09-813-453A-8	Sequence 8, Appl
16	375.5	28.4	246	9 US-09-813-453A-9	Sequence 9, Appl
17	352.5	26.7	212	9 US-09-813-453A-59	Sequence 59, Appl
18	342.5	25.9	273	9 US-09-813-453A-10	Sequence 10, Appl
19	305	23.1	257	9 US-09-813-453A-53	Sequence 53, Appl

20	256.5	19.4	262	9 US-09-813-453A-11	Sequence 11, Appl
21	210.5	15.9	244	9 US-09-813-453A-41	Sequence 41, Appl
22	178.5	13.5	241	9 US-09-813-453A-63	Sequence 63, Appl
23	157	11.9	249	9 US-09-813-453A-70	Sequence 70, Appl
24	150	11.4	257	9 US-09-813-453A-13	Sequence 13, Appl
25	148.5	11.2	229	9 US-09-813-453A-12	Sequence 12, Appl
26	146	11.1	592	9 US-09-813-453A-22	Sequence 22, Appl
27	146	11.1	592	9 US-09-813-453A-43	Sequence 43, Appl
28	145	11.0	460	9 US-09-813-453A-39	Sequence 39, Appl
29	140.5	10.6	267	9 US-09-813-453A-15	Sequence 15, Appl
30	139	10.5	249	9 US-09-813-453A-61	Sequence 61, Appl
31	123.5	9.3	248	9 US-09-813-453A-20	Sequence 20, Appl
32	115	8.7	242	9 US-09-813-453A-65	Sequence 65, Appl
33	87.5	6.6	223	9 US-09-895-913A-74	Sequence 74, Appl
34	87.5	6.6	223	9 US-09-813-453A-14	Sequence 14, Appl
35	87.5	6.6	223	9 US-09-813-453A-67	Sequence 67, Appl
36	84	6.4	225	9 US-09-738-626-4158	Sequence 4158, Ap
37	84	6.4	417	10 US-09-828-313-31	Sequence 31, Appl
38	81.5	6.2	258	9 US-09-738-626-5791	Sequence 5791, Ap
39	81	6.1	610	10 US-09-815-242-11183	Sequence 11183, A
40	81	6.1	1019	1 US-08-834-705-18	Sequence 18, Appl
41	80.5	6.1	380	10 US-09-815-242-10609	Sequence 10609, A
42	80.5	6.1	1864	9 US-09-832-292-27	Sequence 27, Appl
43	80.5	6.1	1885	10 US-09-920-346-2	Sequence 2, Appl
44	80	6.1	732	9 US-10-008-355-6	Sequence 6, Appl
45	79.5	6.0	435	10 US-09-815-242-12025	Sequence 12025, A

ALIGNMENTS

RESULT 1  
US-09-813-453A-49  
; Sequence 49, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn ver. 2.0  
; SEQ ID NO 49  
; LENGTH: 258  
; TYPE: PRT  
; ORGANISM: Bacillus stearothermophilus  
US-09-813-453A-49

Query Match	100.0%;	Score 1321;	DB 9;	Length 258;
Best Local Similarity	100.0%;	Pred. No. 4.3e-126;		
Matches 258;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MIFVLDVGNNTVLGVYDGDDELKHHWRIETSRKTEDEYGMKALLNHVGLQFSDIRGI	60	
Db	1	MIFVLDVGNNTVLGVYDGDDELKHHWRIETSRKTEDEYGMKALLNHVGLQFSDIRGI	60	
QY	61	IISVVPPIMFALERMCLKYFHIKPLIVGPGIKTGLDIKYNPREVGADRIVNAVAGIHL	120	
Db	61	IISVVPPIMFALERMCLKYFHIKPLIVGPGIKTGLDIKYNPREVGADRIVNAVAGIHL	120	
QY	121	YGSPLIIVDFGTFATTCYCYINEHKQYMGGAIAPIGIMISTEALFARAAKLPRIEIRPDDII	180	
Db	121	YGSPLIIVDFGTFATTCYCYINEHKQYMGGAIAPIGIMISTEALFARAAKLPRIEIRPDDII	180	
QY	181	GKNTVSAMQAGILYGVVQVEGIVSRMKAKSKIPPKVIATGGLAPLASESDIIDVDPF	240	

Db 181 GKNTVSAMQAGILYGYVQGVGIVSRMKAKSKIPPVKVIATGGLAPLIASESDIIDVDPF 240  
QY 241 LTLTGLKLLYEKNTKKG 258  
Db 241 LTLTGLKLLYEKNTKKG 258

RESULT 2  
US-09-813-453A-45  
; Sequence 45, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 45  
; LENGTH: 262  
; TYPE: PRT  
; ORGANISM: Bacillus anthracis  
US-09-813-453A-45

Query Match 80.1%; Score 1057.5; DB 9; Length 262;  
Best Local Similarity 77.0%; Pred. No. 2.5e-99;  
Matches 201; Conservative 26; Mismatches 31; Indels 3; Gaps 1;  
QY 1 MIFVLDVGNNTVTLGVYDGDDELKHHWRIETSRKTEDEYGMKALLNHVGLQFSDIRGI 60  
Db 1 MIFVLDVGNNTVTLGVYDGDDELKHHWRIETSRKTEDEYGMKALLNHVGLQFSDIRGI 60  
QY 61 IISVVPPIMFALERMCCKYFHIKPLIVGPGIKTGLDIKYDNPREVGAADRIVNAVAGIHL 120  
Db 61 IVSSVPPIMFALERMCCKYFHIKPLIVGPGIKTGLNIKENPREVGAADRIVNAVAGIHL 120  
QY 121 YGSPLIIVDFGTATTCYINEHKQYMGGAIAPIGIMISTEALFARAAKLPRIEIAIPDDII 180  
Db 121 YGSPLIIVDFGTATTCYINEHKYMGGVITPIMISAEALYSRAAKLPRIEITKPSVV 180  
QY 181 GKNTVSAMQAGILYGYVQGVGIVSRMKAKSKIPPVKVIATGGLAPLIASESDIIDVDPF 240  
Db 181 GKNTVSAMQAGILYGYVQGVGIVSRMKAKSKIPPVKVIATGGLAPLIASESNVIDVDPF 240  
QY 241 LTLTGLKLLYEKNT---EKKG 258  
Db 241 LTLKGLYMLYERNANLQHEKG 261

RESULT 3  
US-09-813-453A-2  
; Sequence 2, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 258  
; TYPE: PRT  
; ORGANISM: Bacillus subtilis  
US-09-813-453A-2  
Query Match 79.2%; Score 1046; DB 9; Length 258;  
Best Local Similarity 78.3%; Pred. No. 3.5e-98;  
Matches 198; Conservative 32; Mismatches 23; Indels 0; Gaps 0;  
QY 1 MIFVLDVGNNTVTLGVYDGDDELKHHWRIETSRKTEDEYGMKALLNHVGLQFSDIRGI 60  
Db 1 LLLVIDVGNNTVTLGVYDGDDELKHHWRIETSRKTEDEYGMKALLNHVGLQFSDIRGI 60  
QY 61 IISVVPPIMFALERMCCKYFHIKPLIVGPGIKTGLDIKYDNPREVGAADRIVNAVAGIHL 120  
Db 61 IISVVPPIMFALERMCCKYFHIKPLIVGPGIKTGLNIKENPREVGAADRIVNAVAGIHL 120  
QY 121 YGSPLIIVDFGTATTCYINEHKQYMGGAIAPIGIMISTEALFARAAKLPRIEIAIPDDII 180  
Db 121 YGNPLIIVDFGTATTCYIDENKQYMGGAIAPIGIMISTEALYSRAAKLPRIEITRPDNI 180  
QY 181 GKNTVSAMQAGILYGYVQGVGIVSRMKAKSKIPPVKVIATGGLAPLIASESDIIDVDPF 240  
Db 181 GKNTVSAMQAGILYGYVQGVGIVSRMKAKSKIPPVKVIATGGLAPLIASESDIIDVDPF 240  
QY 241 LTLTGLKLLYEKN 253  
Db 241 LTLKGLELIYERN 253

RESULT 4  
US-09-813-453A-47  
; Sequence 47, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 47  
; LENGTH: 254  
; TYPE: PRT  
; ORGANISM: Bacillus halodurans  
US-09-813-453A-47

Query Match 73.8%; Score 975; DB 9; Length 254;  
Best Local Similarity 70.8%; Pred. No. 5.6e-91;  
Matches 179; Conservative 36; Mismatches 38; Indels 0; Gaps 0;  
QY 1 MIFVLDVGNNTVTLGVYDGDDELKHHWRIETSRKTEDEYGMKALLNHVGLQFSDIRGI 60  
Db 1 MILVIDVGNNTVTLGVYDQDETLLVHHWRLATSRQKTEDEYAMTVRSFLDHAGLQFQDIDGI 60  
QY 61 IISVVPPIMFALERMCCKYFHIKPLIVGPGIKTGLDIKYDNPREVGAADRIVNAVAGIHL 120  
Db 61 VISSVPPIMFSLQCMCKYFHVTPMIIGPGIKTGLNIKENPREVGAADRIVNAVAAIEL 120  
QY 121 YGSPLIIVDFGTATTCYINEHKQYMGGAIAPIGIMISTEALFARAAKLPRIEIAIPDDII 180  
Db 121 YGYPALIVDFGTATTCYCLINEKKQYAGGVIAPIGIMISTEALYHRASKLPRIEIAKQV 180  
QY 181 GKNTVSAMQAGILYGYVQGVGIVSRMKAKSKIPPVKVIATGGLAPLIASESDIIDVDPF 240

```
Db 181 GTNTIDSMQSGIFGYVQSDGVVKKRKAQAESEPKVIATGGGLAKLIGTESETIDVIDSF 240
QY 241 LTLTGLKLLYEKN 253
Db 241 LTLKGLQLIYKKN 253

RESULT 5
US-09-813-453A-17
; Sequence 17, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 17
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-813-453A-17
```

```
Query Match 68.1%; Score 900; DB 9; Length 233;
Best Local Similarity 78.2%; Pred. No. 2e-83;
Matches 169; Conservative 27; Mismatches 20; Indels 0; Gaps 0;

QY 1 MIFVLDVGNNTNTVLGVYDGDDELKHHWRIETSRSKTEDEYGMKALLNHVGLQFSDIRGI 60
Db 1 MLLVIDVGNNTNTVLGVYHDGKLEYHWRIETSRHKTDEFEFGMLRSLFDHSGLMFEQIDGI 60

QY 61 IISVVPPIMFALERMCIFYHFKPLIVGPGIKTGLDIKYDNPREVGDRIVNAVAGIHL 120
Db 61 IISVVPPIMFALERMCIFYHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120

QY 121 YGSPLIIVDFGTATTTCYINEHKQYMGGAIAPIGIMISTEALFARAAKLPRIEIAIARPDII 180
Db 121 YGNPLIIVDFGTATTTCYIDENKQYMGGAIAPIGITISTEALYSRAAKLPRIETRPDNI 180

Y 181 GKNTVSAMQAGILYGVGVQVEGIVSRMKAISKIPPK 216
Db 181 GKNTVSAMQSGILFGYGVQVEGIVKRMKWQAKQDPR 216
```

```
RESULT 6
US-09-813-453A-7
; Sequence 7, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 7
; LENGTH: 255
```

```
; TYPE: PRT
; ORGANISM: Geobacter sulfurreducens
US-09-813-453A-7

Query Match 63.1%; Score 834; DB 9; Length 255;
Best Local Similarity 62.7%; Pred. No. 1.1e-76;
Matches 160; Conservative 40; Mismatches 55; Indels 0; Gaps 0;

QY 1 MIFVLDVGNNTNTVLGVYDGDDELKHHWRIETSRSKTEDEYGMKALLNHVGLQFSDIRGI 60
Db 1 MLLVIDVGNNTNTVLGIYDGERILVRDWRVSTDKARTTDEYGILINELFRLAGLDQIRAV 60

QY 61 IISVVPPIMFALERMCIFYHFKPLIVGPGIKTGLDIKYDNPREVGDRIVNAVAGIHL 120
Db 61 IISVVPPILTGVLRLSLGVFGMRPLVVGPGIKTGMPIQYDNPREVGDRIVNAVAGYEK 120

QY 121 YGSPLIIVDFGTATTTCYINEHKQYMGGAIAPIGIMISTEALFARAAKLPRIEIAIARPDII 180
Db 121 YRTSLIIVDFGTATTTFDYVNRKGEYCGGAIAPGLVISTEALPQRASKLPDVDIIRPSAII 180

QY 181 GKNTVSAMQAGILYGVGVQVEGIVSRMKAISKIPPKVIATGGGLAPLIASESDIIDVDPF 240
Db 181 ARNTVNSMQAGIYYGVVGLVDEIVTRMKAESKDAPRVATGGLASLIAPESKTIEAVEEY 240

QY 241 LTLTGLKLLYEKNTE 255
Db 241 LTLEGLRILYERNRE 255
```

```
RESULT 7
US-09-813-453A-55
; Sequence 55, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 55
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-09-813-453A-55
```

```
Query Match 61.9%; Score 817.5; DB 9; Length 256;
Best Local Similarity 62.9%; Pred. No. 5.4e-75;
Matches 161; Conservative 40; Mismatches 54; Indels 1; Gaps 1;

QY 1 MIFVLDVGNNTNTVLGVYDGDDELKHHWRIETSRSKTEDEYGMKALLNHVGLQFSDIRGI 60
Db 1 MLLVFDVGNNTNMVLGIYKDGKLVNWRIKTDREKTSDEYGILISNLFYDYNVISDIDDV 60

QY 61 IISVVPPIMFALERMCIFYHFKPLIVGPGIKTGLDIKYDNPREVGDRIVNAVAGIHL 120
Db 61 IISVVPNVMHSLNFICYCKKQPLIVGPGIKTGLNIKYDNPKEVGADRIVNAVAGIEK 120

QY 121 YGSPLIIVDFGTATTTCYINEHKQYMGGAIAPIGIMISTEALFARAAKLPRIEIAIARPDII 180
Db 121 YGAPSLIVDFGTATTTCFAISEKGEYLGGTIAPGIKISSEALFQSASKLPVELAKPGMTI 180

QY 181 GKNTVSAMQAGILYGVGVQVEGIVSRMKAISKIPPK-KVIATGGGLAPLIASESDIIDVDP 239
Db 181 CKSTVSAMQSGIYYGVVGLVVDKIISIMKELNCDDVKVIATGGLAKLIASETSIDYVDG 240
```





[illegible]

US-09-712-363-276  
; Sequence 276, Application US/09712363  
; Patent No. US20020164588A1  
; GENERAL INFORMATION:  
; APPLICANT: Eisenberg, David  
; APPLICANT: Rotstein, Sergio H.  
; APPLICANT: Marcotte, Edward M.  
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND  
; INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS  
; FILE REFERENCE: 07419-032001  
; CURRENT APPLICATION NUMBER: US/09/712,363  
; CURRENT FILING DATE: 2000-11-13  
; PRIOR APPLICATION NUMBER: PCT/US00/02246  
; PRIOR FILING DATE: 2000-01-28  
; PRIOR APPLICATION NUMBER: 60/179,531  
; PRIOR FILING DATE: 2000-02-01  
; PRIOR APPLICATION NUMBER: 60/117,844  
; PRIOR FILING DATE: 1999-01-29  
; PRIOR APPLICATION NUMBER: 60/118,206,  
; PRIOR FILING DATE: 1999-02-01  
; PRIOR APPLICATION NUMBER: 60/126,593  
; PRIOR FILING DATE: 1999-03-26  
; PRIOR APPLICATION NUMBER: 60/134,093.  
; PRIOR FILING DATE: 1999-05-14  
; PRIOR APPLICATION NUMBER: 60/134,092  
; PRIOR FILING DATE: 1999-05-14  
; PRIOR APPLICATION NUMBER: 60/165,124  
; PRIOR FILING DATE: 1999-11-12  
; PRIOR APPLICATION NUMBER: 60/165,086  
; PRIOR FILING DATE: 1999-11-12  
; NUMBER OF SEQ ID NOS: 292  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 276  
; LENGTH: 272  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-09-712-363-276

Query Match 36.0%; Score 475.5; DB 9; Length 272;  
Best Local Similarity 37.5%; Pred. No. 3.1e-40;  
Matches 100; Conservative 57; Mismatches 93; Indels 17; Gaps 5;  
QY 1 MIFVLDVGNNTNTVLGVYDGDDELKHH-----WRIETSRSKTEDEYGMIMKALLNHVGLQF 54  
Db 1 VLLAIDVRNTHTVVGLLSG--MKEHAKVVQOWRIETSEVTADELALTDGL---IGEDS 55  
QY 55 SDIRGIIISVVPPIMFALERMCLKYFHIKP-LIVGPGIKTGLDIKYDNPREVGAIRVN 113  
Db 56 ERLTGTAALSTVPSVLHEVRIMLDQWPSVPHVLTIEPGVRTGIPLLVDPNKEVGADIRVN 115  
114 AVAGIHLGSPLIIVDFGTATTTCYINEHKQYMGGAIAPIGIMISTEALFARAACLPRIEI 173  
Db 116 CLAAAYDRFRKAAIIVDFGSSICVDVWSAKGEFLGGAIAPIGVQVSSDAAARSAAALRRVEL 175  
QY 174 ARPDIIIGKNTVSAMQAGILYGYVQVQVEGIVSRMKA-----KSKIPPVKVIATGGGLAPLIA 228  
Db 176 ARPRSVVGKNTVECMQAGAVFGFAGLVGDGLVGRIREVDVSGFSVDHVAIVATGHTAPLLL 235  
QY 229 SESDIIDVVDPFLLTLGLKLLYEKNTE 255  
Db 236 PELHTVDHYDQHLTLQGLRLVFERNLE 262

RESULT 15  
US-09-813-453A-8  
; Sequence 8, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001

; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 262  
; TYPE: PRT  
; ORGANISM: Deinococcus radiopugnans  
US-09-813-453A-8  
Query Match 31.5%; Score 415.5; DB 9; Length 262;  
Best Local Similarity 35.2%; Pred. No. 3.6e-34;  
Matches 94; Conservative 49; Mismatches 99; Indels 25; Gaps 6;  
QY 2 IFVLDVGNNTNTVLGVYDGD-ELKHHWRIETSRSKTEDEYGMIMKALLNHVGLQFSDIRGI 60  
Db 6 LLAVDIGNTTTTLGLADASGALTHTWRTNRMLPDDLALQLHGLFTLAGAPIP--RAA 63  
QY 61 IISVVVPPI-----MFALERMCLKYFHIKPLIVGPGIKTGLDIKYDNPREVGAIRVNAVA 116  
Db 64 VLSSVAPPVGENYALALKR---HFMDAFAVSAENLPDVTVELDTPGSGVADRLCN--- 116  
QY 117 GIHLYGSP-----LIIVDFGTATTTCYINEHKQYMGGAIAPIGIMISTEALFARAACL 168  
Db 117 ---LFGAEKYLGGLDYAVVVDFTSTNFDVVGRGRFLGGILATGAQVSADALFARAACL 173  
QY 169 PRIETARPDDIIIGKNTVSAMQAGILYGYVQVQVEGIVSRMKAISKIPPKVIATGGGLAPLIA 228  
Db 174 PRITLQAPETAIGKNTVHALQSLVFGYAEMVDGLLRIRRAELPGEAVAVATGGFSRTVQ 233  
QY 229 SESDIIDVVDPFLLTLGLKLLYEKNTE 255  
Db 234 GICQEIDYDETLTLGLVGLVWASRSE 260

Search completed: June 24, 2003, 22:29:02  
Job time : 15.3382 secs



Db 48 DHAGLQFQDIDGIVISSVPPMMFSLQCMCKKYFHVTPMIIGPGIKTGLNIKYDNPKEVG 107  
QY 104 ADRLALCAWSRHLFSEKPVIAVDIGTAITFDVLTGVNRYRGLIMPGLMDMAGALHSRTA 163  
Db 108 ADRIVNAVAIELYG-YPAIVVDFGTATTYCLINEKKQYAGGVIAPGIMISTEALYHRAS 166  
QY 164 QLPQVRIDRPESLLGRSTTECIKSGVFWGVVVKQIGGLVDAIRGDLVRDFGESTVEVIVTG 223  
Db 167 KLPRIEIAKPQVVGNTIDSMQSGIFGYVSQVDGVVKRMKAQ-----AESEPKVIATG 221  
QY 224 GNSRIIVPEIGPVSVIDEAVLRGSDLLLRNMN 256  
Db 222 GLAKLIGTESETIDVIDSFLTLKGLQLIYKKNV 254

RESULT 2

Q9F985 PRELIMINARY; PRT; 258 AA.  
AC Q9F985;  
QY 01-MAR-2001 (TrEMBLrel. 16, Created)  
J 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Putative 32 kDa replication protein.  
OS Bacillus stearothermophilus.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Geobacillus.  
OX NCBI\_TaxID=1422;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=V;  
RA Vasquez C., Pichuantes S., Saavedra C.;  
RL Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF198621; AAG28531.1; -  
DR InterPro; IPR004619; Baf;  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMS; TIGR00671; baf; 1.  
SQ SEQUENCE 258 AA; 28101 MW; 507B55D695095855 CRC64;

Query Match 23.6%; Score 307; DB 2; Length 258;  
Best Local Similarity 28.3%; Pred. No. 1.3e-19;  
Matches 75; Conservative 53; Mismatches 117; Indels 20; Gaps 5;

QY 1 MRLVVDIGNTSTTLAIFTGDE-----EPSVESVPSALFADSSSTMREVFGNMARKHGE 53  
Db 1 MIFVLDVGNNTVLGVYDGDDELKHHWRIETSRGKTEDEY---GMTIKALLNHVGLQFS 57  
QY 54 QAIAICSVVPSATAVGSALLESFSPVLTICCKLRFPRLDYATPHTFGADRLALCAWS 113  
Db 58 DGIITSSVPPIMFALERMCLKPYHFKPIIVGPGIKTGLNIKYDNPREVGAIRVNAVAG 117  
QY 114 RHLFESEKPVIAVDIGTAITFDVLTGVNRYRGLIMPGLMDMAGALHSRTAQLPQVRIDRP 173  
Db 118 IHLVG-SPLIIVDFGTATTYCYINEHKQYMGGAIAPIGIMISTEALFARAACKLPRIE 176  
QY 174 ESLGRSTTECIKSGVFWGVVVKQIGGLVDAIRGDLVRDFGESV--EVIVTGGNSRIIVP 231  
Db 177 DDIIKNTVSAMQAGILYGVGVGEIVSRMK-----AKSPVPPKVIATGGLASLIAS 229  
QY 232 EIGPVSVIDEAVLRGSDLLLRNMN 256  
Db 230 ESNVIDIVDPFLLTLGLKILYKKNV 254

RESULT 3

Q92F54 PRELIMINARY; PRT; 259 AA.  
AC Q92F54;  
QY 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical protein lin0253.  
GN LIN0253.

OS Listeria innocua.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Listeriaceae; Listeria.  
OX NCBI\_TaxID=1642;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CLIP 11262 / SEROVAR 6A;  
RX PubMed=11679669;  
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
RA Baquero F., Bloecker P., Bloecker H., Brandt P., Chakraborty T.,  
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,  
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,  
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,  
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,  
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;  
RT "Comparative genomics of Listeria species.";  
RL Science 294:849-852(2001).  
DR EMBL; AL596164; CAC95486.1; -  
DR ListiList; LIN00253; -  
DR InterPro; IPR004619; Baf;  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMS; TIGR00671; baf; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 259 AA; 28227 MW; 554B03A0C0EFA64F CRC64;

Query Match 23.6%; Score 307; DB 16; Length 259;  
Best Local Similarity 30.0%; Pred. No. 1.3e-19;  
Matches 82; Conservative 50; Mismatches 107; Indels 34; Gaps 7;

QY 1 MRLVVDIGNTSTTLAIF-----TGDEEPSVESVPSALFADSSSTMREVFGNMARK 49  
Db 1 MILVIDVGNTNCTGVYKEQKLLRHRWMTDRHRTSDEL-----GMTVLNFFSYANLT 53  
QY 50 HGEPQAIACSVVPSATAVGSALLESFSPVLTICCKLRFPRLDYATPHTFGADRL-- 107  
Db 54 PSDIQIISVPPIMHAMETMCVRYFNIRPLIVGPGIKTGLNKLKVDNPREIGSDRIVN 113  
QY 108 ALCAWSRHLFSEK---PVIADVDTAITFDVLTGVNRYRGLIMPGLMDMAGALHSRTAQ 164  
Db 114 AVAA-----SEYGTVPVIVDFGTATTFCYIDEAGVYQGGAIAPIGIMISTEALYNRAK 167  
QY 165 LPQVRIDRPESLLGRSTTECIKSGVFWGVVVKQIGGLVDAIRGDLVRDFGESTVEVIVTG 224  
Db 168 LPRVDIAESSQIIGKSTVASMQAQIFGYFIGQCEGII----AEMKQSNTPV-VVATGG 222  
QY 225 NSRIIVPEIGPVSVIDEAVLRGSDLLLRNMNMP 257  
Db 223 LARMIITEKSSAVDILDPPFLTLKGLELLYRRNKP 255

RESULT 4

Q8YAC5 PRELIMINARY; PRT; 259 AA.  
ID Q8YAC5  
AC Q8YAC5;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical protein lmo0221.  
GN LMO0221.

OS Listeria monocytogenes.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Listeriaceae; Listeria.  
OX NCBI\_TaxID=1639;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=EGD-E / SEROVAR 1/2A;  
RX MEDLINE=21537279; PubMed=11679669;  
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
RA Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,



RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,  
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,  
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
RA Jones L.-M., Kaerst U., Krefit J., Kuhn M., Kunst F., Kurapkat G.,  
RA Madueno E., Maltournam A., Mata Vicente J., Ng E., Nedjari H.,  
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,  
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,  
RT \*Comparative genomics of *Listeria species*.  
RL Science 294:849-852(2001).  
DR EMBL; AL591974; CAD00748.1; -  
DR MEROPS; M41.009; -  
DR ListList; LMO00221; -  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMS; TIGR00671; baf; 1.  
KW Hypothetical protein; Complete proteome.  
Q SEQUENCE 259 AA; 28187 MW; F106B049D80025B7 CRC64;  
Query Match 23.6%; Score 307; DB 16; Length 259;  
Best Local Similarity 30.4%; Pred. No. 1.3e-19;  
Matches 83; Conservative 49; Mismatches 107; Indels 34; Gaps 7;  
QY 1 MRLVVDIGNTSTTLAIF-----TGDEEPSVESVPSALFADSDSTMREVFNGMARK 49  
Db 1 MILVIDVGNNTCTVGVYKQKLLKHWRMTDRHRTSDEL-----GMTVLNFFSYANLT 53  
QY 50 HGEPOAIAICSVVPSATAVGSALLESLFVSVPLTICCKLRFPRLDYATPHTFGADRL-- 107  
Db 54 PSDIQIISVVPPIMHAMETMCVRYFNIRPLIVGPIKTLNKLVDNPREIGSDRIVN 113  
QY 108 ALCAWSRHLFSEK---PVIADVIGTAITFDVLTGNYRGGLIMPGIDMMAGALHSRTAQ 164  
Db 114 AVAA-----SEEGTPTVIVVDEGTATTCYIDESGVYGGAIAPGIMISTEALYNRAK 167  
QY 165 LPQVRIDRPSLLGRSTECIKSGVFWGVVKQIGGLVDAIRGDLVRDFGESTVEIVTGG 224  
Db 168 LPRVDIAESSQIIGKSVSSMQAGIFYGVGQCEGII---AEMKKQSNASPV-VVATGG 222  
QY 225 NSRIIVPEIGPVSVIDEAVLRGSDLLLRNMMP 257  
Db 223 LARMITEKSSAVDILDPLTLKGLLELLYRRNKP 255  
RESULT 5  
Q8XHL5 PRELIMINARY; PRT; 259 AA.  
AC Q8XHL5;  
DT 01-MAR-2002 (TremBLrel. 20, Created)  
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TremBLrel. 21, last annotation update)  
DE Hypothetical protein CPE2468.  
GN CPE2468.  
OS Clostridium perfringens.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
OC Clostridiales; Clostridiaceae; Clostridium.  
OX NCBI\_TaxID=1502;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=13 / TYPE A;  
RX PubMed=11792842;  
RA Shmizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,  
RA Shiba T., Ogawara N., Hattori M., Kuhara S., Hayashi H.,  
RT \*Complete genome sequence of *Clostridium perfringens*, an anaerobic  
RT flesh-eater.  
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).  
DR EMBL; AP003194; BAB82174.1; -  
DR InterPro; IPR004619; Baf.  
DR InterPro; IPR000515; BPD\_transp.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMS; TIGR00671; baf; 1.  
DR PROSITE; PS00402; BPD\_TRANSP\_INN\_MEMBR; UNKNOWN\_1.

KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 259 AA; 28819 MW; OD5FEA3B7A145E10 CRC64;  
Query Match 23.2%; Score 302; DB 16; Length 259;  
Best Local Similarity 28.1%; Pred. No. 3.6e-19;  
Matches 74; Conservative 62; Mismatches 113; Indels 14; Gaps 6;  
QY 1 MRLVVDIGNTSTTLAIFTGDEEPSVE----SVPSALFAD--SSTMREVFNGMARKHGEPO 54  
Db 1 MILLIDVGNNTNIVLGIH--DNEKYIASWRISTDSKKTSDSEYSIQVMQLENOAKLPEDVE 58  
QY 55 AIAICSVVPSATAVGSALLESLFVSVPLTICCKLRFPRLDYATPHTFGADRL--ALCAW 112  
Db 59 GIIISVVVPNIMHSLNEMVRKCFCKEPIVVGPIKGTINIKYDNPKEVGDRIVNAVAAF 118  
QY 113 SRHLESEKPVIAVDIGTAITFDVLTGNYRGGLIMPGIDMMAGALHSRTAQLPQVRIDR 172  
Db 119 EKH---KKPMIIDFGTATTCAITERKGYLGGNICPGIQISADALFERAAKLPRIELEK 175  
QY 173 PESLLGRSTTECIKSGVFWGVVKQIGGLVDAIRGDLVRDFGESTVEIVTGNRSRIIVPE 232  
Db 176 PKSVICKNTVTSMQAGIIGYIGKVEYIVKRMKKEMM-DLGEKEPFVLATGTLAKLVYSE 234  
QY 233 IGPVSVIDEAVLRGSDLLLRMN 255  
Db 235 TDVIDEVDRLTLEGLKILYEKN 257  
RESULT 6  
Q9X8N6 PRELIMINARY; PRT; 265 AA.  
AC Q9X8N6;  
DT 01-NOV-1999 (TremBLrel. 12, Created)  
DT 01-NOV-1999 (TremBLrel. 12, Last sequence update)  
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)  
DE Hypothetical protein SCO3380.  
GN SCO3380 OR SCE94.31C.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Oliver K., Harris D.,  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.,  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RX MEDLINE=97000351; PubMed=8843436;  
RA Redenbach M., Kieser H.M., Denapalite D., Eichner A., Cullum J.,  
RA Kinashi H., Hopwood D.A.,  
RT "A set of ordered cosmids and a detailed genetic and physical map for  
RT the 8 Mb *Streptomyces coelicolor* A3(2) chromosome."  
RL Mol. Microbiol. 21:77-96(1996).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2) / M145;  
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,  
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,  
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
RA Hopwood D.A.,  
RT \*Complete genome sequence of the model actinomycete *Streptomyces*

RT coellicolor A3(2).";  
RL Nature 417:141-147(2002).  
DR EMBL; AL049628; CAB40880.1; -  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMS; TIGR00671; baf; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 265 AA; 28203 MW; E34FA2ABBA946969 CRC64;  
  
Query Match 22.6%; Score 295; DB 16; Length 265;  
Best Local Similarity 32.9%; Pred. No. 1.6e-18;  
Matches 94; Conservative 42; Mismatches 96; Indels 54; Gaps 13;  
  
QY 1 MRLVVDIGNTSTTLAIFTGDEEPSVE---SVPSALFADST-----MREVFNGMAR 48  
Db 1 MLTTIDVGNTHTVLGLFDG--EDIVEHWRISTDSRRTADELAVLLQGLMGHPLLGD--- 55  
  
QY 49 KHGEP-QAIAICSVVPSAT-----AVGSALLES--LFSVPVLTICCKLRPFPR 93  
Db 56 ELGDGIDGIAICATVPSVLHRELVTRRYGDPVAVLVEPGVKTGPILT----- 105  
/ 94 LDYATPHTFGADRLALCAWSRHLFSEKPVIAVDIGTAITFDVLDTVGNYRGGGLIMPIDM 153  
Db 106 -DH--PKEVGADRIINAVAVELYG-GPAIVVDFGATFDVARSARGEYIGGVIAPGIEI 161  
  
QY 154 MAGALHSRTAQLPQVRIDRPESLLGRSTTECIKSGVFWGVVKQIGGLVDAIRGDLVRDFG 213  
Db 162 SVEALGVKGAQLRKIEVARPSVIGKNTVEAMQSGIYVGFAGQVDGVNRMARELADD-- 219  
  
QY 214 ESTVEVITGNSRIIPEIGPVSVIDE---LAVLRGSDLLLRNM 256  
Db 220 PDDVTVIATGGLAPMV---LGESSVIDEHEPWLTLMLGLRLVYERNV 262

RESULT 7

Q8RFE4  
ID Q8RFE4 PRELIMINARY; PRT; 256 AA.  
AC Q8RFE4;  
DT 01-JUN-2002 (TremBLrel. 21, Created)  
DT 01-JUN-2002 (TremBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)  
DE Bvg accessory factor.  
GN FN0761.  
OS Fusobacterium nucleatum (subsp. nucleatum).  
OC Bacteria; Fusobacteria; Fusobacterium.  
OX NCBI\_TaxID=76856;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 25586;  
MEDLINE=21886394; PubMed=11889109;  
Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,  
Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,  
Vasileva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,  
Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,  
Fonstein M., Kyrpides N., Overbeek R.;  
RT "Genome sequence and analysis of the oral bacterium Fusobacterium  
nucleatum strain ATCC 25586."  
RL J. Bacteriol. 184:2005-2018(2002).  
DR EMBL; AE010586; AAL94957.1; -  
KW Complete proteome.  
SQ SEQUENCE 256 AA; 28601 MW; CDF681127F08914B CRC64;  
  
Query Match 22.1%; Score 288; DB 16; Length 256;  
Best Local Similarity 31.6%; Pred. No. 6.4e-18;  
Matches 84; Conservative 43; Mismatches 117; Indels 22; Gaps 7;  
  
QY 1 MRLVVDIGNTSTTLAIFTGDEEPSVESVPSALFADSTMR-----VFGNMARKHGE--- 52  
Db 1 MIIGIDIGNTH----IVTGIYDNNGELISTFRATNDKMTEDYEYFSYFNNT-KYNEISI 55  
  
QY 53 --PQAIACSVVPSATAVGSALLESLSFVPLTICCKLRFPRLDYATPHT-FGADRLAL 109  
Db 56 KKVDAILLISSVVPNIITTFQFFARKYFKVEATIVDLKKLPFTFAKGINVTGFGADRIID 115

QY 110 CAWSRHLFSEKPVIAVDIGTAITFDVLDTVGNYRGGGLIMPIDMAGALHSRTAQLPQVR 169  
Db 116 ITEAMOKYPDKNLVIFDFGTATTYDVLKK-GVYIGGGILPGIDMSINALYGNATKLPRVK 174  
  
QY 170 IDRPESLGRSTTECIKSGVFWGVVKQIGGLVDAIRGDLVRDFGSESTVEVITGNSRII 229  
Db 175 FTTPSSVLGTDTKQIQAAIFFGYAGQIKHIKKINEEL-----NEEIFVLATGGLGKIL 229  
  
QY 230 VPEIGPVSVIDEAVLRGSDLLLRMN 255  
Db 230 SAEIDEIDEYDANLSLKGLTYLYKLN 255  
  
RESULT 8  
Q9CD56  
ID Q9CD56 PRELIMINARY; PRT; 274 AA.  
AC Q9CD56;  
DT 01-JUN-2001 (TremBLrel. 17, Created)  
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)  
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)  
DE Hypothetical protein ML0232.  
GN ML0232.  
OS Mycobacterium leprae.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1769;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-TN;  
RX MEDLINE=21128732; PubMed=11234002;  
RA Cole S.T., Biglmeier K., Parkhill J., James K.D., Thomson N.R.,  
Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,  
Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,  
Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
Barrell B.G.;  
RT "Massive gene decay in the leprosy bacillus."  
RL Nature 409:1007-1011(2001).  
DR EMBL; AL583917; CAC29740.1; -  
DR Leproma; ML0232; -  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMS; TIGR00671; baf; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 274 AA; 29421 MW; 1C2E735BDEC78765 CRC64;

Query Match 21.5%; Score 280.5; DB 16; Length 274;  
Best Local Similarity 29.8%; Pred. No. 3.3e-17;  
Matches 79; Conservative 53; Mismatches 120; Indels 13; Gaps 7;  
  
QY 1 MRLVVDIGNTSTTLAIFTGDEEPS--VES----VPSALFADSTMRVFGNMARKHGEPO 54  
Db 1 MLLAIDVRNTHTVVGLSGSKEHAKVQWRIQRTSEVTADELAL--IIDGLIGDDSERL 58  
  
QY 55 A-IAICSVVPSATAVGSALLESLF-SVPVLTICCKLRFPRLDYATPHTFGADRLALCAW 112  
Db 59 AGAAALSTVPSVLHEVRIMLDQYWPSPVPHVLEPGVRTGIPLLVDPNPKVAGADRVNCLA 118  
  
QY 113 SRHLESEKPVIAVDIGTAITFDVLDTVGNYRGGGLIMPIDMAGALHSRTAQLPQVRIDR 172  
Db 119 AFHKGQAAIV-VDFGSSICVDVVSAGKEFLGGAIPGVQVSSDAAAARSALRRVELAR 177  
  
QY 173 PESLLGRSTTECIKSGVFWGVVKQIGGLVDAIRGDLVRDFG--STVEVITGNSRIIV 230  
Db 178 PRSVGKNTVECMQAGVVVGFAGLVGDMRQDVVEEFGSLGNRVAVVATGHTAPLLL 237  
  
QY 231 PEIGPVSVIDEAVLRGSDLLLRMN 255  
Db 238 PELHTVDHYDRHLTLHGLRLVVERN 262









**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run On: June 24, 2003, 21:49:47 ; Search time 10.4272 Seconds  
(without alignments)  
725.188 Million cell updates/sec

Title: US-09-813-453A-53  
Perfect score: 1303  
Sequence: 1 MRLVVDIGNTSTTLAIFTGD.....VIDELAVLRGSDLLLRMNP 257

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5.

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	98	7.5	782	4 US-09-543-084A-32	Sequence 32, Appl
2	98	7.5	782	4 US-09-543-084A-33	Sequence 33, Appl
3	98	7.5	782	4 US-09-543-084A-36	Sequence 36, Appl
4	96	7.4	782	4 US-09-543-084A-34	Sequence 34, Appl
5	96	7.4	782	4 US-09-543-084A-35	Sequence 35, Appl
6	82	6.3	297	2 US-09-027-013-3	Sequence 3, Appl
7	82	6.3	297	3 US-09-244-233-3	Sequence 3, Appl
8	80.5	6.2	491	4 US-09-362-899-3	Sequence 3, Appl
9	79.5	6.1	774	1 US-08-019-870-3	Sequence 3, Appl
10	79	6.1	431	3 US-08-478-507-2	Sequence 2, Appl
11	79	6.1	431	4 US-09-128-275A-2	Sequence 2, Appl
12	79	6.1	431	4 US-09-553-427-2	Sequence 2, Appl
13	79	6.1	991	4 US-09-877-730-12	Sequence 12, Appl
14	79	6.1	1069	4 US-09-877-730-2	Sequence 2, Appl
15	79	6.1	1072	4 US-09-877-730-18	Sequence 18, Appl
16	79	6.1	1150	4 US-09-877-730-8	Sequence 8, Appl
17	79	6.1	1693	3 US-08-478-507-7	Sequence 7, Appl
18	79	6.1	1693	4 US-09-128-275A-7	Sequence 7, Appl
19	79	6.1	1693	4 US-09-553-427-7	Sequence 7, Appl
20	78.5	6.0	533	2 US-08-225-488-2	Sequence 2, Appl
21	78.5	6.0	774	1 US-08-019-870-8	Sequence 8, Appl
22	78	6.0	576	2 US-08-676-279-58	Sequence 58, Appl
23	78	6.0	1891	2 US-08-804-227C-12	Sequence 12, Appl
24	78	6.0	1891	2 US-08-804-198-6	Sequence 6, Appl
25	77.5	5.9	430	4 US-09-134-001C-2981	Sequence 2981, Ap
26	77.5	5.9	773	1 US-08-019-870-1	Sequence 1, Appl
27	77.5	5.9	773	1 US-08-019-870-6	Sequence 6, Appl

28	77.5	5.9	774	1 US-07-747-901A-3	Sequence 3, Appl
29	77.5	5.9	774	1 US-07-935-312-3	Sequence 3, Appl
30	77.5	5.9	774	1 US-08-019-870-11	Sequence 11, Appl
31	77.5	5.9	774	1 US-08-633-760-44	Sequence 44, Appl
32	77.5	5.9	774	1 US-08-633-760-46	Sequence 46, Appl
33	77.5	5.9	774	1 US-08-633-760-50	Sequence 50, Appl
34	77.5	5.9	774	1 US-08-633-760-52	Sequence 52, Appl
35	76.5	5.9	609	4 US-08-969-683A-67	Sequence 67, Appl
36	76	5.8	3729	2 US-08-804-227C-4	Sequence 4, Appl
37	75.5	5.8	600	6 5240706-1	Patent No. 5240706
38	75.5	5.8	740	1 US-08-309-512-10	Sequence 10, Appl
39	75.5	5.8	740	5 PCT-US92-08756A-10	Sequence 10, Appl
40	75.5	5.8	774	1 US-08-314-309A-21	Sequence 21, Appl
41	75.5	5.8	1544	4 US-09-413-814-46	Sequence 46, Appl
42	74.5	5.7	652	2 US-08-426-125-9	Sequence 9, Appl
43	74.5	5.7	652	2 US-08-455-355-9	Sequence 9, Appl
44	74.5	5.7	712	4 US-09-877-730-22	Sequence 22, Appl
45	74.5	5.7	774	1 US-08-633-760-48	Sequence 48, Appl

ALIGNMENTS

RESULT 1  
US-09-543-084A-32  
; Sequence 32, Application US/09543084A  
; Patent No. 6361988  
; GENERAL INFORMATION:  
; APPLICANT: Frances H. Arnold  
; APPLICANT: Zhixin Shao  
; APPLICANT: Huimin Zhao  
; APPLICANT: Lorraine J. Giver  
; TITLE OF INVENTION: Recombination of Polynucleotide  
; TITLE OF INVENTION: Sequences Using Defined or Random Primer Sequences  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Oppenheimer Wolff & Donnelly LLP  
; STREET: 2029 Century Park East, Suite 3800  
; CITY: Los Angeles  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 90067  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: Windows 98  
; SOFTWARE: Word 97  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/543,084A  
; FILING DATE: April 4, 2000  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/381,935  
; APPLICATION NUMBER: 60/041,666  
; FILING DATE: March 25, 1997  
; APPLICATION NUMBER: 60/045,211  
; FILING DATE: April 30, 1997  
; APPLICATION NUMBER: 60/046,256  
; FILING DATE: May 12, 1997  
; APPLICATION NUMBER: 08/905,359  
; FILING DATE: August 4, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Oldenkamp, David J.  
; REGISTRATION NUMBER: 29,421  
; REFERENCE/DOCKET NUMBER: 330187-89  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (310) 788-5000  
; TELEFAX: (310) 788-5100  
; INFORMATION FOR SEQ ID NO: 32:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 782 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear

```
; MOLECULE TYPE: peptide
US-09-543-084A-32

Query Match          7.5%; Score 98; DB 4; Length 782;
Best Local Similarity 24.3%; Pred. No. 0.0072;
Matches 65; Conservative 32; Mismatches 75; Indels 96; Gaps 15;

QY 7 IGNTSTTLAIFT--GDEE-----PSVESVPSALFADSSMTREVFNGNMARKHGEPOA---- 55
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 512 LGNERTPRSLRTRGLDQIQORLAGTDGLPGKGTFTARLWQVMFGN--RMHGAELARDDL 569
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 56 IAICSVVPSATAVGSALLESLFSVPVLTICCKL--RPFRLDYATPHTFGADRLALCAWS 113
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 570 VALCRRQPTATASNGAIVD-----LTAAC TALS RFDERAD-----LDSRG 609
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 114 RHLFSEKPVIAVDIGTAITFDVLDTVGNRGGLIMPGIDMMAGALHSRTAQLPQVRIDRP 173
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 610 AHLFTEF-ALAGGIRFADTFEVTDPV-----RT-----P 637
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 174 ESLGRSTTECIKSGVFWGVVKQIGGL-VDAIRGDLVRDF-----GES-TVEV 219
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 638 RRL---NTTDPVRTALADAVQRLAGIPLDAKLGDHITDSRGERRIPIHGGRGEAGTFNV 694
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 220 I-----VTGNSRIIVPEIGP 235
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 695 ITNPLVPGVGYPQVVHGTSTFVMAVELGP 722
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 2
US-09-543-084A-33
; Sequence 33, Application US/09543084A
; Patent No. 6361988
; GENERAL INFORMATION:
; APPLICANT: Frances H. Arnold
; APPLICANT: Zhixin Shao
; APPLICANT: Huimin Zhao
; APPLICANT: Lorraine J. Giver
; TITLE OF INVENTION: Recombination of Polynucleotide
; TITLE OF INVENTION: Sequences Using Defined or Random Primer Sequences
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oppenheimer Wolff & Donnelly LLP
; STREET: 2029 Century Park East, Suite 3800
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90067
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 98
; SOFTWARE: Word 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/543,084A
; FILING DATE: April 4, 2000
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/381,935
; APPLICATION NUMBER: 60/041,666
; FILING DATE: March 25, 1997
; APPLICATION NUMBER: 60/045,211
; FILING DATE: April 30, 1997
; APPLICATION NUMBER: 60/046,256
; FILING DATE: May 12, 1997
; APPLICATION NUMBER: 08/905,359
; FILING DATE: August 4, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Oldenkamp, David J.
; REGISTRATION NUMBER: 29,421
; REFERENCE/DOCKET NUMBER: 330187-89
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (310) 788-5000
; TELEFAX: (310) 788-5100
```

```
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 782 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-543-084A-33

Query Match          7.5%; Score 98; DB 4; Length 782;
Best Local Similarity 24.3%; Pred. No. 0.0072;
Matches 65; Conservative 32; Mismatches 75; Indels 96; Gaps 15;

QY 7 IGNTSTTLAIFT--GDEE-----PSVESVPSALFADSSMTREVFNGNMARKHGEPOA---- 55
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 512 LGNERTPRSLRTRGLDQIQORLAGTDGLPGKGTFTARLWQVMFGN--RMHGAELARDDL 569
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 56 IAICSVVPSATAVGSALLESLFSVPVLTICCKL--RPFRLDYATPHTFGADRLALCAWS 113
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 570 VALCRRQPTATASNGAIVD-----LTAAC TALS RFDERAD-----LDSRG 609
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 114 RHLFSEKPVIAVDIGTAITFDVLDTVGNRGGLIMPGIDMMAGALHSRTAQLPQVRIDRP 173
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 610 AHLFTEF-ALAGGIRFADTFEVTDPV-----RT-----P 637
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 174 ESLGRSTTECIKSGVFWGVVKQIGGL-VDAIRGDLVRDF-----GES-TVEV 219
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 638 RRL---NTTDPVRTALADAVQRLAGIPLDAKLGDHITDSRGERRIPIHGGRGEAGTFNV 694
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 220 I-----VTGNSRIIVPEIGP 235
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 695 ITNPLVPGVGYPQVVHGTSTFVMAVELGP 722
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 3
US-09-543-084A-36
; Sequence 36, Application US/09543084A
; Patent No. 6361988
; GENERAL INFORMATION:
; APPLICANT: Frances H. Arnold
; APPLICANT: Zhixin Shao
; APPLICANT: Huimin Zhao
; APPLICANT: Lorraine J. Giver
; TITLE OF INVENTION: Recombination of Polynucleotide
; TITLE OF INVENTION: Sequences Using Defined or Random Primer Sequences
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oppenheimer Wolff & Donnelly LLP
; STREET: 2029 Century Park East, Suite 3800
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90067
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 98
; SOFTWARE: Word 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/543,084A
; FILING DATE: April 4, 2000
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/381,935
; APPLICATION NUMBER: 60/041,666
; FILING DATE: March 25, 1997
; APPLICATION NUMBER: 60/045,211
; FILING DATE: April 30, 1997
; APPLICATION NUMBER: 60/046,256
; FILING DATE: May 12, 1997
; APPLICATION NUMBER: 08/905,359
; FILING DATE: August 4, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Oldenkamp, David J.
```



```

; REGISTRATION NUMBER: 29,421
; REFERENCE/DOCKET NUMBER: 330187-89
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (310) 788-5000
; TELEFAX: (310) 788-5100
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 782 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
;
US-09-543-084A:36

```

Query Match	7.5%;	Score 98;	DB 4;	Length 782;
Best Local Similarity	24.3%;	Pred. No. 0.0072;		
Matches 65;	Conservative 32;	Mismatches 75;	Indels 96;	Gaps 15;

QY	56	IAICSVPSATAVGSALLESLFSVPVLTI	CKKL--RFPRLDYATPHTFGADRLA	CAWS	113
Dd	570	VALCRRQPTATASNGAIVD-----LTAAC	TALSRFDERAD-----LD	SRG	609
QY	114	RHLFSEKPVIAVDIGTITFDVLDTVGN	RGGLIMP GIDMMAGALHSRTAQLPQVR	IDRP	173
Dd	610	AHLFTFEF-ALAGGIRFADTFEVTDPV-	-----RT-----P		637
QY	174	ESLLGRSTTECIKSGVFVGWVKQIGGL-VDA	IRGDLVRDF-----GES-TVEV		219
Dd	638	RRL---NTTDPVRTALADAVQRLAGIPLD	AKLGDIHTDSRGERRIPIHGGRGEAGT	FNV	694

QY 220 I-----VTGGSRSRIIVPEIGP 235  
| | | | : :  
Db 695 ITNPLVPGVGYPQVVHGTSFVNMAVELGP 722

RESULT 4  
US-09-543-084A-34  
; Sequence 34, Application US/09543084A  
; Patent No. 6361988  
; GENERAL INFORMATION:  
; APPLICANT: Frances H. Arnold  
; APPLICANT: Zhixin Shao  
; APPLICANT: Huimin Zhao  
; APPLICANT: Lorraine J. Giver  
TITLE OF INVENTION: Recombination of Polynucleotide  
TITLE OF INVENTION: Sequences Using Defined or Random Primer Sequences

FILING DATE: May 12, 1997  
 APPLICATION NUMBER: 08/905,359  
 FILING DATE: August 4, 1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Oldenkamp, David J.  
 REGISTRATION NUMBER: 29,421  
 REFERENCE/DOCKET NUMBER: 330187-89  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (310) 788-5000  
 TELEFAX: (310) 788-5100  
 INFORMATION FOR SEQ ID NO: 34:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 782 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide

Query Match	7.4%	Score 96;	DB 4;	Length 782;
Best Local Similarity	23.9%;	Pred. No. 0.013;		
Matches 64;	Conservative 32;	Mismatches 76;	Indels 96;	Gaps 15;
QY	7	IGNTSTTLAIFT--GDEE----	PSVESVPSALFADSSMTREVFGNMARKHG----	EPQA 55
DB	512	LGNERTPRSLRFLGDLQIQORLAGTDGLPGKGTARLWQVMFGN--	RMHGAELVRDDL 569	
QY	56	IAICSVVPSATAVGSALLESLFSVPVLTICCKL--RFPFRLDYATPHTFGADRLALCAWS	113	
DB	570	VALCRRQPTATASNGAIVD-----LTAACLTALSRFDERAD-----	LDSRG 609	
QY	114	RHLFSEKPVIAVDIGTAITFDVLDTVGNYRGGLIMPGLMDMAGALHSRTAQLPQVRIDRP	173	
DB	610	AHLFTEF--ALAGGIRFADTFEVTDPV-----	RT-----P 637	
QY	174	ESLLGRSTTECIKSGVFWGVVKQIGGL--VDAIRGDLVRDF-----	GES-TVEV 219	
DB	638	RRL---NTTDPVRVTALADAVQRLAGIPLDAKLGLDIHTDSRGERRIPIHGGRGAGTFNV	694	
QY	220	I-----VTGGNSRIIVPEIGP	235	
DB	635	ITNPLPVGVGIPQVVHGTSFVMAVELGP	722	

RESULT 5  
US-09-543-084A-35  
; Sequence 35, Application US/09543084A  
; Patent No. 6361988  
; GENERAL INFORMATION:  
; APPLICANT: Frances H. Arnold  
; APPLICANT: Zhixin Shao  
; APPLICANT: Huimin Zhao  
; APPLICANT: Lorraine J. Giver  
; TITLE OF INVENTION: Recombination of Polynucleotide  
; TITLE OF INVENTION: Sequences Using Defined or Random Primer Sequences  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Oppenheimer Wolff & Donnelly LLP  
; STREET: 2029 Century Park East, Suite 3800  
; CITY: Los Angeles  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 90067  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: Windows 98  
; SOFTWARE: Word 97  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/543,084A  
; FILING DATE: April 4, 2000  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/381,935

APPLICATION NUMBER: 60/041,666  
FILING DATE: March 25, 1997  
APPLICATION NUMBER: 60/045,211  
FILING DATE: April 30, 1997  
APPLICATION NUMBER: 60/046,256  
FILING DATE: May 12, 1997  
APPLICATION NUMBER: 08/905,359  
FILING DATE: August 4, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Oldenkamp, David J.  
REGISTRATION NUMBER: 29,421  
REFERENCE/DOCKET NUMBER: 330187-89  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (310) 788-5000  
TELEFAX: (310) 788-5100  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 782 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
;-09-543-084A-35

Query Match 7.4%; Score 96; DB 4; Length 782;  
Best Local Similarity 23.9%; Pred. No. 0.013;  
Matches 64; Conservative 32; Mismatches 76; Indels 96; Gaps 15;  
QY 7 IGNTSTTLAIFT--GDEE-----PSVESVPSALFADSSSTMREVFGNMARKHG-----EPQA 55  
Db 512 LGNERTPRSLRTRGLDQIQORLAGTDGLPGKGTARLWQVMFGN--RMHGAELVRDDL 569  
QY 56 IAICSVVPSATAVGSALLESLSFVPLTICCKL--REFRLDYATPHTFGADRLALCAWS 113  
Db 570 VALCRRQPTATASNGAIVD-----LTAACALSRFDERAD-----LDSRG 609  
QY 114 RHLFSEKPVIAVDIGTAITFDVLTGVNRYRGGLIMPIDMAGALHSRTAQLPQVRIDRP 173  
Db 610 AHLTFEF-ALAGGIRFADTFEVTDPV-----RT-----P 637  
QY 174 ESSLGRSTTECIKSGVFWGVVKQIGGL-VDAIRGDLVRDF-----GES-TVEV 219  
Db 638 RRL---NTDPRVRTALADAVORLAGIPLDALKLDIHTDSRGERRIPIHGGRGEAGTFNV 694  
QY 220 I-----VTGGSNRIIVPEIGP 235  
Db 695 ITNPLVPGVGYPQVVHGTSEVMVELGP 722

RESULT 6  
S-09-027-013-3  
Sequence 3, Application US/09027013  
Patent No. 5962302  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Shah, Purvi  
APPLICANT: Corley, Neil C.  
TITLE OF INVENTION: HUMAN N-ACETYLNEURAMINATE LYASE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/027,013  
FILING DATE: Herewith

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0462 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 297 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: GI 42131  
US-09-027-013-3

Query Match 6.3%; Score 82; DB 2; Length 297;  
Best Local Similarity 22.2%; Pred. No. 0.14;  
Matches 55; Conservative 34; Mismatches 91; Indels 68; Gaps 12;  
QY 33 FADSSSTMREVFGNMARKHGEPOAIAICSVVPSATAVGSALLES--FSVPVLTICCKLRF 90  
Db 52 FVQSLSEREQVLEIVAEEGKIKLIAHVGCVTTAESQQLAASAKRYGFDVAVSAVTFYY 111  
QY 91 PFRLDYATPH---TFGADRLALCAWSRHLFSEKPVIAVDI---GTAITFDVLDTVGNY 142  
Db 112 PFSFEHCHYRAIIDSADGL-----PMVVYNIPALSGVKLTLDQINTL--- 155  
QY 143 RGGLIMPIDMAGALHSRTA---QLPQVRIDRPESLLGRSTTECIKSGVFWGVVKQIG- 198  
Db 156 ---VTLPGV---GALKQTSGLDLYQMEQIRREHPDLVLYNGYDEIFASGLLAGADGGIGS 208  
QY 199 -----GLVDAIR-GDLVRDFGESTVEVITGGSNRIIVPEIGPVSVIDEL---AV 244  
Db 209 TYNIMGWRYQGIVKALKEGDI-----QTAQKLQTECN-----KVIDLLIKTGV 251  
QY 245 LRGSLLLL 252  
Db 252 FRGLKTVL 259

RESULT 7  
US-09-244-233-3  
Sequence 3, Application US/09244233  
Patent No. 6030824  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Shah, Purvi  
APPLICANT: Corley, Neil C.  
TITLE OF INVENTION: HUMAN N-ACETYLNEURAMINATE LYASE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/244,233  
FILING DATE:  
CLASSIFICATION:

; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 09/027,013  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Billings, Lucy J  
 ; REGISTRATION NUMBER: 36,749  
 ; REFERENCE/DOCKET NUMBER: PF-0462 US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 650-855-0555  
 ; TELEFAX: 650-845-4166  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 297 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; IMMEDIATE SOURCE:  
 ; LIBRARY: GenBank  
 ; CLONE: GI 42131  
 ; JS-09-244-233-3

	Query Match	6.3%;	Score 82;	DB 3;	Length 297;
	Best Local Similarity	22.2%;	Pred. No. 0.14;		
	Matches	55;	Conservative 34;	Mismatches 91;	Indels 68;
					Gaps 12;
QY	33	FADSSTMREYFGNMARKHGE	QPAIAICSVVPSATAVGS	SALLESL--FSVPVLTICCKLRF	90
			: : :	: : :	:
Db	52	FVQSLSEREOVLEIVAEEGK	GKIKLIAHVGCVTTAESQQLA	ASAKRYGFDVAVSAVTPFY	111
			: : :	: : :	:
QY	91	PERLDYATPH----	TFGADRLALCAWSRHLFSEK	PIAVDI-----GTAITFDVLDTV	142
			: : :	: : :	:
Db	112	PFSFEECHDYRAIDSADGL	-----PMVVYNIPALSGVKLT	LDQINTL---	155
			: : :	: : :	:
QY	143	RGLLIMPGIDMMAGALHSRTA	---QLPQVRIDRPESLLGRST	TECIKSGVFWGVVKIG-	198
		: :	: :       :	: :       :	:
Db	156	---VTLPGV-----	GALKQTSGLYQMEQIRREHP	DLVLYNGYDEIFASGLLAGADGG	208
			: : :	: : :	:
QY	199	-----GLVDAIR-GDLVR	DGFESTVEVIVTGGNSRIIV	PEIGPVSVIDEL----	244
			: : :	: : :	:
Db	209	TYNIMGWRYOGIVKALKEGDI	-----QTAQKLQTECN-	-----KVIDLLIKTV	251
			: : :	: : :	:
QY	245	LRGSDLLL	252		
			: : :	: : :	:
Db	252	FRGLKTVL	259		
			: : :	: : :	:

## RESULT 8

```

;S-09-362-899-3
; Sequence 3, Application US/09362899
; Patent No. 6361986
; GENERAL INFORMATION:
; APPLICANT: Degussa-Hls AG
; APPLICANT: Forschungszentrum-Jlich
; TITLE OF INVENTION: PROCESS FOR TH
; CURRENT APPLICATION NUMBER: US/09/
; CURRENT FILING DATE: 1999-07-29
; EARLIER APPLICATION NUMBER: DE 199
; EARLIER FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 491
; TYPE: PRT
; ORGANISM: Corynebacterium glutami
US-09-362-899-3

```

	Query Match	6.2%; Score 80.5; DB 4;	Length 491;
	Best Local Similarity	21.5%; Pred. No.	0.48;
	Matches 64; Conservative	47; Mismatches 111;	Indels 75; Gaps 16;
OY	15 AIFTGDEEPSVESVPSSA-LEADSSSTMREVFGNMARKHCEPQAIACISVWPVSATAVGSALL	73	
Dd	62 SVITGG--EGTVEGIPVAIVILSDFSFLGGSLGTVA-----SVRIMKAIHRATELKLPLL	112	

QY	74	ESLFS-----VPVLTICCKLR-----FPFRLDYATPHTFGADRLALCAW--S	113
Db	113	VSPASGGARMQEDNRAFYVMVSITAAVQVRHREAHLPF-LVYLRNPMTMG--AMASWGSS	168
QY	114	RHLFSEKP-----VIAVDIGTAITFDVLDTVGNRYGGGLIMPIDMMAGALHSRTA	163
Db	169	GHLTFAEPGAQIGFLPRVVELTTGHALPDGVQQAENLVKTGV----IDGIVSPLQLRAA	224
QY	164	QLPQVRIDRPESLLGR--STTECIKSGVFWGVVKQ-----IGGLVDAIRGDLVRDFGE	214
Db	225	VAKTLKVIQPVAEATDRFSPPTFGVALPVMEAIARSRDPQRPQGEIMEITLGADVVKLSGA	284
QY	215	ST-----VEVIVTGGNSRIIVPE-----IGPVSVIDEA-VLRGSDLLLRNMNP	257
Db	285	RAGALSPAVRVALARIGRPVVVLIGQDRRFTLGP-----QELRFARRGISLARELNLP	337

## RESULT 9

US-08-019-870-3  
; Sequence 3, Application US/08019870  
; Patent No. 5336613  
; GENERAL INFORMATION:  
; APPLICANT: NIWA, MINEO  
; APPLICANT: YOSHIMASA, SAITO  
; APPLICANT: SASAKI, HITOSHI  
; APPLICANT: ISHII, YOSHINORI  
; TITLE OF INVENTION: A NEW CEPHALOSPORIN C ACYLASE  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
; ADDRESSEE: P.C.  
; STREET: 1755 S. Jefferson Davis Highway, Suite 400  
; City: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/019,870  
; FILING DATE: 19930219  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Oblon, No. 5336613man F.  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 18-791-0  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 413-3000  
; TELEFAX: (703) 413-2220  
; TELEX: 248855 OPAT UR  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 774 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-019-870-3

Query Match 6.1%; Score 79.5; DB 1; Length 774;  
Best Local Similarity 23.4%; Pred. No. 1.3;  
Matches 62; Conservative 29; Mismatches 101; Indels 73; Gaps 13;

```

QY      1  MRLVVDIGNSTTTLAIFTGDEEPSVESVPSEALFADSSMTREVFNGMARKHGEPQAI-- 58
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     187  LKLYDDGGRD-LLCIPGAERDRLE-----ADLATLRPAVDALLKAMGGDASDAAG 238

QY      59  ---CSVVPSEATAVGSALL-----ESLFSVP-----VLTICKLRFPF-- 92
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     239  GSNNAVAPGRTATGRPILAGDPHRVFEIPGMYAQHHLACDRFDMTGLTVPGVPGFPHFA 298

```









GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run On: June 24, 2003, 22:24:12 ; Search time 14.2826 Seconds  
(without alignments)  
1947.059 Million cell updates/sec

Title: US-09-813-453A-53  
Perfect score: 1303  
Sequence: 1 MRLVVDIGNTSTTLAIFTGD.....VIDELAVLRGSDLLLRMNP 257

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues  
Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1303	100.0	257	9 US-09-813-453A-53	Sequence 53, Appl
2	322	24.7	254	9 US-09-813-453A-47	Sequence 47, Appl
3	319	24.5	258	9 US-09-813-453A-2	Sequence 2, Appl
4	316	24.3	262	9 US-09-813-453A-45	Sequence 45, Appl
5	305	23.4	258	9 US-09-813-453A-49	Sequence 49, Appl
6	301	23.1	255	9 US-09-813-453A-7	Sequence 7, Appl
7	295	22.6	265	9 US-09-813-453A-4	Sequence 4, Appl
8	274.5	21.1	272	9 US-09-813-453A-5	Sequence 5, Appl
9	273.5	21.0	233	9 US-09-813-453A-17	Sequence 17, Appl
10	271.5	20.8	272	9 US-09-712-363-276	Sequence 276, App
11	265.5	20.4	219	9 US-09-813-453A-57	Sequence 57, Appl
12	265	20.3	258	9 US-09-813-453A-6	Sequence 6, Appl
13	261	20.0	260	9 US-09-813-453A-51	Sequence 51, Appl
14	254.5	19.5	256	9 US-09-813-453A-55	Sequence 55, Appl
15	253.5	19.5	262	9 US-09-813-453A-8	Sequence 8, Appl
16	229.5	17.6	250	9 US-09-813-453A-3	Sequence 3, Appl
17	228.5	17.5	246	9 US-09-813-453A-9	Sequence 9, Appl
18	227	17.4	241	9 US-09-813-453A-63	Sequence 63, Appl
19	225	17.3	229	9 US-09-813-453A-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1  
US-09-813-453A-53  
; Sequence 53, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 53  
; LENGTH: 257  
; TYPE: PRT  
; ORGANISM: Chlorobium tepidum  
US-09-813-453A-53

Query Match	100.0%;	Score 1303;	DB 9;	Length 257;
Best Local Similarity	100.0%;	Pred. No. 1.7e-126;		
Matches 257;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MRLVVDIGNTSTTLAIFTGDEEPSVESVPSALFADSTMTREVFGNMARKHGEPOAIAICS	60	
Db	1	MRLVVDIGNTSTTLAIFTGDEEPSVESVPSALFADSTMTREVFGNMARKHGEPOAIAICS	60	
QY	61	VVPSATAVGSALLESLFSPVLTICCKLRFPFRLDYATPHTFGADRLALCAWSRHLFSEK	120	
Db	61	VVPSATAVGSALLESLFSPVLTICCKLRFPFRLDYATPHTFGADRLALCAWSRHLFSEK	120	
QY	121	PVIAVDIGTAITFDVLDTVGNRYRGLIMPIDMAGALHSRTAQLPQVRIDRPESLLGRS	180	
Db	121	PVIAVDIGTAITFDVLDTVGNRYRGLIMPIDMAGALHSRTAQLPQVRIDRPESLLGRS	180	
QY	181	TTECIKSGVFWGVVKQIGGLVDAIRGDLVRDFGESTVEVIVTGGNSRIIVPEIGPVSVID	240	

```
Db 181 TTECIKSGVFWGVVVKQIGGLVDAIRGDLVRDFGSESTVEVIVTGGNSRIIVPEIGPVSVID 240
QY 241 ELAVLRGSDLLLRNMNP 257
    |||||
Db 241 ELAVLRGSDLLLRNMNP 257

RESULT 2
US-09-813-453A-47
; Sequence 47, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Bacillus halodurans
US-09-813-453A-47
```

```
Query Match 24.7%; Score 322; DB 9; Length 254;
Best Local Similarity 28.9%; Pred. No. 3.3e-25;
Matches 79; Conservative 54; Mismatches 104; Indels 36; Gaps 4;

QY 1 MRLVVDIGNTSTTLAIF-----TGDEEPSVESVPSALFADSSSTMREVFG 43
    |||:||||:|:|:|
Db 1 MILVIDGNTNTVLGVYQDET|LVHHWRLATSRQKTEDEY-----AMTVRS|F 47
    : : : | | | | |

QY 44 GNMARHGEPQAIACSVVPSATAVGSALLESLFVSVPLTICCKLRFPFRLDYATPHTFG 103
    : : : | | | | |
Db 48 DHAGLQFQDIDGIVISSVPPMMFSLQEMCKRYFHVTPMIIGPGIKTGLN|KYNPKEVG 107

QY 104 ADRLCAWSRHLFSEKPVIAVDIGTAITFDVLDTVGNRGGLIMPGIDMMAGALHSRTA 163
    |||: : : | | | | | : : : | | | | | : : : | | : : : |
Db 108 ADRIVNAVAAILYGY-YP|AVVDFGTATTYCLINEKKQYAGGVIAPGIMISTEALYHRAS 166

QY 164 QLPQVRIDRPESLLGRSTTECIKSGVFWGVVVKQIGGLVDAIRGDLVRDFGSESTVEVIVTG 223
    |||: : : | : : : | : : : | : : : | : : : | : : : | : : : |
b 167 KLPRIEIAKPKQVVGNTIDSMQSGIFGYVSQDGVVVKRMKAQ-----AESEPKVIATG 221

QY 224 GNSRIIVPEIGPVSVIDE|LAVLRGSDLLLRNMN 256
    | : : | | : | | | | | | : : | : : | : : | : : |
Db 222 GLAKLIGTESETIDVIDSFLTLKGLQLIYKKNV 254
```

```
RESULT 3
US-09-813-453A-2
; Sequence 2, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
```

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-813-453A-2

Query Match 24.5%; Score 319; DB 9; Length 258;
Best Local Similarity 28.1%; Pred. No. 6.8e-25;
Matches 76; Conservative 52; Mismatches 106; Indels 36; Gaps 4;

QY 3 LVVDIGNTSTTLAIF-----TGDEEPSVESVPSALFADSSSTMREVFGN 45
    |||:||||:|:|:|
Db 3 LVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMI-----LRS|FDH 49
    : : : | | | | |

QY 46 MARKHGEPQAIACSVVPSATAVGSALLESLFVSVPLTICCKLRFPFRLDYATPHTFGAD 105
    : : : | | | | |
Db 50 SGLMFEQIDGII|SSVPP|PFALERMCTKYFHIEPQIVGPGMKTGLN|KYNPKEVGAD 109

QY 106 RLALCAWSRHLFSEKPVIAVDIGTAITFDVLDTVGNRGGLIMPGIDMMAGALHSRTAQL 165
    | : : | | | | | : : : | | | | | : : : | | : : : |
Db 110 RIVNAVAAILYGY-NPL|VVDFGTATTYCYIDENKQYMGGA|APGITISTEALYSRAAKL 168

QY 166 PQVRIDRPESLLGRSTTECIKSGVFWGVVVKQIGGLVDAIRGDLVRDFGSESTVEVIVTG 225
    ||:| : ||:| : ||:| : ||:| : ||:| : ||:| : ||:| : ||:| : ||:|
Db 169 PRIETRPDNIIGKNTVSAMQSGILFGYVGQVEGIVKRMKWOAKOD-----LKVIATGGL 223

QY 226 SRIIVPEIGPVSVIDE|LAVLRGSDLLLRMN 255
    : : | : : | : : | : : | : : | : : | : : | : : |
Db 224 APLIANESDCIDIVDPF|TLKGLLELIYERN 253
```

```
RESULT 4
US-09-813-453A-45
; Sequence 45, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
; LENGTH: 262
; TYPE: PRT
; ORGANISM: Bacillus anthracis
US-09-813-453A-45
```

```
Query Match 24.3%; Score 316; DB 9; Length 262;
Best Local Similarity 28.9%; Pred. No. 1.4e-24;
Matches 76; Conservative 56; Mismatches 113; Indels 18; Gaps 5;

QY 1 MRLVVDIGNTSTTLAIF|TGDEEPSVESVPSALFADSSSTMREVFGNMARK-----HGE 52
    | : : | | : | : | | : | : | : | : | : | : | : | : | : | : |
Db 1 MIFVLDVGNTNAVLGVF---EEGELRQ-HWRMETDRHKTEDEYGM|LVKQLEHGLSFED 56
    : : : | | | | | : : : | : : : | : : : | : : : | : : : |

QY 53 PQAIACSVVPSATAVGSALLESLFVSVPLTICCKLRFPFRLDYATPHTFGADRLALCAW 112
    : : : | | | | | : : : | : : : | : : : | : : : | : : : |
Db 57 VKGIIVSSVPP|PFALERMCEKYFK|KPLVVGP|G|K|TGLN|KYNP|PREVGADRIVNAVA 116

QY 113 SRHLFSEKPVIAVDIGTAITFDVLDTVGNRGGLIMPGIDMMAGALHSRTAQLPQVRIDR 172
    ||:| : ||:| : ||:| : ||:| : ||:| : ||:| : ||:| : ||:| : ||:|
Db 117 GIHLYG-SPL|I|VDFGTATTYCYINEEKHYMG|GVITP|G|M|S|A|E|A|L|Y|S|R|A|A|K|L|P|R|I|E|I|T|K 175

QY 173 PESLLGRSTTECIKSGVFWGVVVKQIGGLVDAIRGDLVRDFGSESTVEVIVTGGNSRIIVPE 232
```



Db 176 PSSVVGKNTVSAMQSGILYGVQVEGIVKRMKEE-----AKQEPKVIATGGLAKLISEE 230  
QY 233 IGPVSVIDELAVLRGSDLLLRMN 255  
Db 231 SNVIDVVDPPFLTKGLKGLMYLYERN 253

RESULT 5  
US-09-813-453A-49  
; Sequence 49, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 49  
; LENGTH: 258  
; TYPE: PRT  
; ORGANISM: Bacillus stearothermophilus  
US-09-813-453A-49

Query Match 23.4%; Score 305; DB 9; Length 258;  
Best Local Similarity 29.3%; Pred. No. 1.9e-23;  
Matches 76; Conservative 52; Mismatches 121; Indels 10; Gaps 4;  
QY 1 MRLVVDIGNTSTTLAIFTGDE---EPSVESVPSALFADSSM-REVFGNMARKHGEPOAI 56  
Db 1 MIFVLDVGNNTNVLGVGDDELKHHWRIETSRKTEDEYGMKALLNHVGLQFSDIRGI 60  
QY 57 AICSVVPSATAVGSALLESLFSVPVLTICCKLRFPRLDYATPHTFGADRLALCAWSRHL 116  
Db 61 IISVVPIMFALERMCKLYFHFKPLIVGPGIKTGDKIDYDNPREVGADRIVNAVAGIHL 120  
QY 117 FSEKPVIAVDIGTAITFDVLDTVGNRGGLIMPIDMAGALHSRTAQLPQVRIDRPESL 176  
Db 121 YG-SPLIIVDFGTATCYINEHKQYMGGAIAPIGIMISTEALFARAACKLPRIEIRPDDI 179  
QY 177 LGRSTTECIKSGVFWGVVKQIGGLVDAIRGDLVRDFGESTVEVITGGSRIIVPEIGPV 236  
Db 180 IGKNTVSAMQAGILYGVQVEGIVSRMKAK-----SKIPPKVIATGGLAPLIASESDII 234  
QY 237 SVIDELAVLRGSDLLLRMN 255  
Db 235 DVVDPPFLTLTGLKLLYEKN 253

RESULT 6  
US-09-813-453A-7  
; Sequence 7, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21

; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 255  
; TYPE: PRT  
; ORGANISM: Geobacter sulfurreducens  
US-09-813-453A-7  
Query Match 23.1%; Score 301; DB 9; Length 255;  
Best Local Similarity 31.8%; Pred. No. 4.9e-23;  
Matches 85; Conservative 48; Mismatches 108; Indels 26; Gaps 8;  
QY 1 MRLVVDIGNTSTTLAIFTGDEEPSVESVPSALFADSSMREVFG-----NMARKHG-----E 52  
Db 1 MLLVIDVGNNTNVLGIYDGERLVRDWRVST---DKARTTDEYGILINELFRLAGLGLDQ 56  
QY 53 PQAIAICSVVPSATAVGSALLESLFSVPVLTICCKLRFPRLDYATPHTFGADRL--ALC 110  
Db 57 IRAVIISVVPPLTGVLERLSLGYFGMRPLVVGPGIKTGMPLOYDNPREVGADRIVNAVA 116  
QY 111 AWSRHLFSEKPVIAVDIGTAITFDVLDTVGNRGGLIMPIDMAGALHSRTAQLPQVRI 170  
Db 117 GYEKYRTS---LIIVDFGTATFDYVNRKGEYCGGAIAPGLVISTEALFQRAKSLPRVDI 173  
QY 171 DRPESLLGRSTTECIKSGVFWGVVKQIGGLVDAIRGDLVRDFGES--TVEVITGGSRI 228  
Db 174 IRPSAIIARNTVNSMQAGIYYGVV---GLVDEI---VTRMKAESKDAPRVATGGLASL 226  
QY 229 IVPEIGPVSVIDE LAVLRGSDLLLRMN 255  
Db 227 IAPESKTIEAVEEYLTLEGLRILYERN 253

RESULT 7  
US-09-813-453A-4  
; Sequence 4, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 265  
; TYPE: PRT  
; ORGANISM: Streptomyces coelicolor  
US-09-813-453A-4

Query Match 22.6%; Score 295; DB 9; Length 265;  
Best Local Similarity 32.9%; Pred. No. 2.1e-22;  
Matches 94; Conservative 42; Mismatches 96; Indels 54; Gaps 13;  
QY 1 MRLVVDIGNTSTTLAIFTGDEEPSVE---SVPSALFADSSST-----MREVFNGMAR 48  
Db 1 MLLTIDVGNTHTVLGLFDG--EDIVEHWRISTDSRRTADELAVLLQGLMGHMLLGD--- 55  
QY 49 KHGEP-QAIAICSVVPSAT-----AVGSALLES--LFSVPVLTICCKLRFPFR 93  
Db 56 ELGDGIDGIAICATVPSVLHRELRETVRRYGDVPAVLVEPGVKTGPILT----- 105  
QY 94 LDYATPHTFGADRLALCAWSRHLFSEKPVIAVDIGTAITFDVLDTVGNRGGLIMPIDM 153  
Db 106 -DH--PKEVGADRIINAAVAVELYG-GPAIVVDFGTATTDFDAVSARGEYIGVIAPIGIEI 161

QY 154 MAGALHSRTAQLPQVRIDRPESLLGRSTTECIKSGVFWGVVVKQIGGLVDAIRGDLVRDFG 213  
Db 162 SVEALGVKGAQLRKIEVARPSVIGKNTVEAMQSGIVYGAGQVDGVNRMARELADD-- 219  
QY 214 ESTVEVITGCGNSRIIVPEIGPVSVIDE---LAVLRGSDLLLRNM 256  
Db 220 PDDVTVIATGGLAPMV---LGESSVIDEHEPWLTLMLGLRLVYERNV 262

RESULT 8  
US-09-813-453A-5  
; Sequence 5, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 272  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-09-813-453A-5

Query Match 21.1%; Score 274.5; DB 9; Length 272;  
Best Local Similarity 29.3%; Pred. No. 2.9e-20;  
Matches 77; Conservative 52; Mismatches 125; Indels 9; Gaps 5;  
QY 1 MRLVVDIGNSTTLAIFTGDEEPS--VES----VPSALFADSSSTMREVFGNMARKHGEPO 54  
Db 1 MLLAIDVRNTHTVVGLLSGMKEHAKVVQWRIQRTSEVTADELAL-TIDGLIGEDSERLT 59  
QY 55 AIAICSVVPSATAVGSALLESLF-SVPVLTICCKLRFPFRLDYATPHTFGADRLALCAWS 113  
Db 60 GTAALSTVPSVLHEVRIMLDQYWPSVPHVLIEPGVRTGIPLLVDPNPKVGGADRVNCLAA 119  
QY 114 RHLFSEKPVIAVDIGTAITFDVLDTVGNRYGGLIMPIDMAGALHSRTAQLPQVRIDRP 173  
Db 120 YDRF-RKAAIVVDFGSSICVDVVSAGKEFLGGAIAPGVQVSSDAAAARSALRRVELARP 178  
QY 174 ESLGRSTTECIKSGVFWGVVVKQIGGLVDAIRGDLVRDFGSESTVEVITGCGNSRIIVPEI 233  
Db 179 RSVGKNTVECMQAGAVFGFAGLVGDLVGRIRIEDVSGFSVDHDAIVATGHTAPLLLP 238  
QY 234 GPNVVIDELAVLRGSDLLLRNM 256  
Db 239 HTVDHYDQHLTLQGLRLVFERNL 261

RESULT 9  
US-09-813-453A-17  
; Sequence 17, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569

; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 17  
; LENGTH: 233  
; TYPE: PRT  
; ORGANISM: Bacillus subtilis  
US-09-813-453A-17

Query Match 21.0%; Score 273.5; DB 9; Length 233;  
Best Local Similarity 28.5%; Pred. No. 3e-20;  
Matches 65; Conservative 42; Mismatches 90; Indels 31; Gaps 3;  
QY 1 MRLVVDIGNSTTLAIF-----TGDEEPSVESVPSALFADSSSTMREV 43  
Db 1 MLLVIDVGNNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMI-----LRS 47  
QY 44 GNMARKHGEPOAIAICSVVPSATAVGSALLESLFVSVVLTICCKLRFPFRLDYATPHTFG 103  
Db 48 DHSGLMFEQIDGIIISVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVG 107  
QY 104 ADRLALCAWSRHLFSEKPVIAVDIGTAITFDVLDTVGNRYGGLIMPIDMAGALHSRTA 163  
Db 108 ADRIVNAVAIIHLYG-NPLIVVDFTATTTCYIDENKQYMGGAIPGTTISTEALYSRAA 166  
QY 164 QLPQVRIDRPESLLGRSTTECIKSGVFWGVVVKQIGGLVDAIRGDLVRD 211  
Db 167 KLPRIETRPDNIIGKNTVSAMQSGILFGYGVQVEGIVKRMKWAQKOD 214

RESULT 10  
US-09-712-363-276  
; Sequence 276, Application US/09712363  
; Patent No. US20020164588A1  
; GENERAL INFORMATION:  
; APPLICANT: Eisenberg, David  
; APPLICANT: Rotstein, Sergio H.  
; APPLICANT: Marcotte, Edward M.  
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND  
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS  
; FILE REFERENCE: 07419-032001  
; CURRENT APPLICATION NUMBER: US/09/712,363  
; CURRENT FILING DATE: 2000-11-13  
; PRIOR APPLICATION NUMBER: PCT/US00/02246  
; PRIOR FILING DATE: 2000-01-28  
; PRIOR APPLICATION NUMBER: 60/179,531  
; PRIOR FILING DATE: 2000-02-01  
; PRIOR APPLICATION NUMBER: 60/117,844  
; PRIOR FILING DATE: 1999-01-29  
; PRIOR APPLICATION NUMBER: 60/118,206,  
; PRIOR FILING DATE: 1999-02-01  
; PRIOR APPLICATION NUMBER: 60/126,593  
; PRIOR FILING DATE: 1999-03-26  
; PRIOR APPLICATION NUMBER: 60/134,093  
; PRIOR FILING DATE: 1999-05-14  
; PRIOR APPLICATION NUMBER: 60/134,092  
; PRIOR FILING DATE: 1999-05-14  
; PRIOR APPLICATION NUMBER: 60/165,124  
; PRIOR FILING DATE: 1999-11-12  
; PRIOR APPLICATION NUMBER: 60/165,086  
; PRIOR FILING DATE: 1999-11-12  
; NUMBER OF SEQ ID NOS: 292  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 276  
; LENGTH: 272  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-09-712-363-276

Query Match 20.8%; Score 271.5; DB 9; Length 272;  
Best Local Similarity 29.1%; Pred. No. 5.9e-20;  
Matches 76; Conservative 52; Mismatches 124; Indels 9; Gaps 5;



QY 52 EPQAIACSVVPSATAVGSALLESLFVSVPLTI--CCKLRFPFRLDYATPHTFGADRLAL 109  
Db 57 -IDAVIISVVPSQIFNLRLNLSRRYFNVEPLVIGENAKLIDVRIE--KPSEAGADRLVN 113  
QY 110 CAWSRHLFSEKPVIAVDIGTAITFDVLDTVGNRYRGGGLIMPIDMMAGALHSRTAQLPQVR 169  
Db 114 AIGAAWVY-PGPLVVIDSGTATTFDIVAADGAFEGGIAPGINLSMOALHEAAKLPRIA 172  
QY 170 IDRP--ESLLGRSTTECIKSGVFWGVVVKQIGGLVDAIRGDLVRDFGESTVEVIVTGGNSR 227  
Db 173 IORPAGNRIVGTDFTVSAMQSGVFWGYISLIEGLVARIKAER----GE-PMTVIATGGVAS 227  
QY 228 IIVPEIGPVSVIDEAVLRGSDLLLRMN 255  
Db 228 LFEGATDSIDHFDSDLTIRGLLEIYRRN 255

RESULT 14  
US-09-813-453A-55  
Sequence 55, Application US/09813453A  
Patent No. US20020168681A1  
GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Yocum, R. Rogers  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 55  
; LENGTH: 256  
; TYPE: PRT  
; ORGANISM: Clostridium difficile  
US-09-813-453A-55

Query Match 19.5%; Score 254.5; DB 9; Length 256;  
Best Local Similarity 28.6%; Pred. No. 3.1e-18;  
Matches 80; Conservative 47; Mismatches 102; Indels 51; Gaps 10;  
QY 1 MRLVVDIGNTSTTLAIFTGD-----EPPSVE--SVPSALF-ADSSTMREVFGN 45  
Db 1 MLLVFDVGNTMVLGIYKDKLVNYWRIKTDREKTSDEYGILISLNFYDYNISDI--- 57  
Y 46 MARKHGEPOAIAICSVVPSATAVGSALLESLFVSVPLTICCK-----LRPFRLDYA 97  
Db 58 -----DDVIISVVPN-----VMHSENF-CIKYCKKQPLIVGPGIKTGLNIKYD 101  
QY 98 TPHTFGADRL--ALCAWSRHLFSEKPVIAVDIGTAITFDVLDTVGNRYRGGGLIMPIDMMA 155  
Db 102 NPKQVGADRIVNAVAGIEKY---GAPSVLVDFTATTFCATSEKGEYLGGTIAPGIKISS 158  
QY 156 GALHSRTAQLPQVRIDRPESLLGRSTTECIKSGVFWGVVVKQIGGLVDAIRGDLVRDFGES 215  
Db 159 EALFQSASKLPRVELAKPGMTICKSTVSAMQSGIIGYV---GLVDKIISIMKKELNCD 214  
QY 216 TVEIVITGNSRIIVPEIGPVSVIDEAVLRGSDLLLRMN 255  
Db 215 DVKVIATGGLAKLIASETKSIDYVDGFLTLEGLRIIYEKN 254

RESULT 15  
US-09-813-453A-8  
; Sequence 8, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers

; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 262  
; TYPE: PRT  
; ORGANISM: Deinococcus radiopugnans  
US-09-813-453A-8

Query Match 19.5%; Score 253.5; DB 9; Length 262;  
Best Local Similarity 27.8%; Pred. No. 4.1e-18;  
Matches 74; Conservative 46; Mismatches 103; Indels 43; Gaps 7;  
QY 3 LVVDIGNTSTTLAIFTGTDEEPSVESVPSALFADSS-----TMREVFGN--MARKH 50  
Db 7 LAVDIGNTTTLVGL-----ADASGALHTWIRTNREMLPDDLALQLH 49  
QY 51 G-----EPQAIACSVVPSATAVGSALLESLFVSVPLTICCKLRFPFRLDYATPHTF 102  
Db 50 GLFTLAGAPIPRAAVLSSVAPPVGENYALALKRHFEMIDAFASAEENLPDVTVELDTPGSV 109  
QY 103 GADRLA-LCAWSRHLFSEKPVIAVDIGTAITFDVLDTVGNRYRGGGLIMPIDMMAGALHSR 161  
Db 110 GADRLCNLFGEAEKYLGGIDYAVVDFGTSTNFDVVGRGRRFLGGILATGAQVSADALFAR 169  
QY 162 TAQLPQVRIDRPESLLGRSTTECIKSGVFWGVVVKQIGGLVDAIRGDLVRDFGESTVEVIV 221  
Db 170 AAKLPRIITLQAPETAIGKNTVHALQSGLVFGYAEVMDGLLRIRRAELP---GEAV--AVA 224  
QY 222 TGGNSRIIVPEIGPVSVIDEAVLRG 247  
Db 225 TGGFSRTVQGICQEIDYYDETTLRG 250

Search completed: June 24, 2003, 22:29:04  
Job time: 15.2826 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:46:39 ; Search time 29.091 seconds  
(without alignments)  
1177.181 Million cell updates/sec

Title: US-09-813-453A-53  
Perfect score: 1303  
Sequence: 1 MRLVVDIGNSTTIAFTGD.....VIDELAVLFGSDLLLRMNP 257

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :		A_Geneseq_101002.*	
1:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*		
2:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*		
3:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*		
4:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*		
5:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*		
6:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*		
7:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*		
8:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*		
9:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*		
10:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*		
11:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*		
12:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*		
13:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*		
14:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*		
15:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*		
16:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*		
17:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*		
18:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*		
19:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*		
20:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*		
21:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*		
22:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*		
23:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1303	100.0	257	AAU911174	Pantothenate kinas
2	322	24.7	254	AAU911171	Pantothenate kinas
3	319	24.5	258	AAU01243	B. subtilis novel
4	319	24.5	258	AAU911149	Bacillus subtilis
5	316	24.3	262	AAU911170	Pantothenate kinas
6	307	23.6	259	ABB47661	Listeria monocytog
7	305	23.4	258	AAU911172	Pantothenate kinas
8	301	23.1	255	AAU911154	Geobacter sulfurre
9	295	22.6	265	AAU911151	Streptomyces coeli
10	274.5	21.1	272	AAU911152	Mycobacterium tube

11	273.5	21.0	233	23	AAU911163	Pantothenate kinas
12	271.5	20.8	272	22	AAG81225	Mycobacterium tube
13	265.5	20.4	219	23	AAU911176	Pantothenate kinas
14	265	20.3	258	23	AAU911153	Rhodobacter capsul
15	261	20.0	260	23	AAU911173	Pantothenate kinas
16	254.5	19.5	256	23	AAU911175	Pantothenate kinas
17	253.5	19.5	262	23	AAU911155	Deinococcus radiop
18	229.5	17.6	250	23	AAU911150	Clostridium acetob
19	228.5	17.5	246	23	AAU911156	Thermotoga maritim
20	227	17.4	241	23	AAU911179	Pantothenate kinas
21	225	17.3	229	23	AAU911159	Aquifex aeolicus p
22	221	17.0	244	23	AAU911168	Pantothenate kinas
23	214	16.4	212	23	AAU911177	Pantothenate kinas
24	197.5	15.2	273	23	AAU911157	Treponema pallidum
25	169	13.0	389	21	AAU74909	Neisseria meningit
26	166.5	12.8	262	23	AAU911158	Borrelia burgdorfe
27	162.5	12.5	592	20	AAU38615	Neisseria meningit
28	162.5	12.5	592	21	AAU74912	Neisseria meningit
29	162.5	12.5	592	23	AAU911169	Pantothenate kinas
30	161.5	12.4	455	20	AAU38617	Neisseria gonorrhoe
31	161.5	12.4	455	21	AAU74908	Neisseria gonorrhoe
32	161.5	12.4	460	23	AAU911167	Pantothenate kinas
33	161.5	12.4	592	20	AAU38618	Neisseria gonorrhoe
34	151.5	12.4	592	21	AAU74911	Neisseria gonorrhoe
35	159.5	12.2	455	21	AAU74910	Neisseria meningit
36	159.5	12.2	592	20	AAU38616	Neisseria meningit
37	159.5	12.2	592	21	AAU74913	Neisseria meningit
38	159.5	12.2	592	23	AAU911166	Pantothenate kinas
39	157	12.0	248	23	AAU911164	Pantothenate kinas
40	153.5	11.8	249	23	AAU911182	Pantothenate kinas
41	146.5	11.2	257	23	AAU911160	Synechocystis pant
42	144	11.1	242	23	AAU911180	Pantothenate kinas
43	136.5	10.5	267	23	AAU911162	Bordella pertussis
44	130.5	10.0	249	23	AAU911178	Pantothenate kinas
45	119	9.1	189	20	AAU38614	Neisseria meningit

ALIGNMENTS

RESULT 1	
AAU911174	
ID	AAU911174 standard; Protein; 257 AA.
XX	
AC	AAU911174;
XX	
DT	05-JUN-2002 (first entry)
XX	
DE	Pantothenate kinase (Coax) #12.
XX	
KW	Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW	pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX	
OS	Chlorobium tepidum.
XX	
PN	WO200216601-A2.
XX	
PD	28-FEB-2002.
XX	
PF	24-AUG-2001; 2001WO-US26531.
XX	
PR	24-AUG-2000; 2000US-227860P.
PR	20-MAR-2001; 2001US-0813453.
XX	
PA	(OMNI-) OMNIGENE BIOPRODUCTS INC.
XX	
PI	Yocum RR, Patterson TA;
XX	
DR	WPI; 2002-269358/31.
DR	N-PSDB; ABK54195.
XX	
PT	Identifying potential antibiotic or antimicrobial agent, comprises contacting composition comprising pantothenate kinase (Coax) protein



XX WPI; 2001-218644/22.  
DR N-PSDB; AAS00984.  
XX  
PT New recombinant microorganism which overexpress a Bacillus subtilis  
PT pantothenate biosynthetic enzyme, useful for the high yield production  
PT of panto-compounds such as pantothenate and pantoate -  
XX  
PS Example 14; Fig 23; 292pp; English.  
XX  
CC The sequence represents a novel B. subtilis pantothenate kinase (encoded  
CC by gene coaX), an enzyme of the pantothenate biosynthetic pathway.  
CC Pantothenate, also known as vitamin B5, is used as a nutritional  
CC supplement in mammals and humans. The invention concerns methods of  
CC producing recombinant microorganisms overexpressing at least one Bacillus  
CC subtilis pantothenate biosynthetic enzyme. The microorganisms and methods  
CC of producing them are useful for producing a panto-compound such as  
CC pantothenate or pantoate, which is a nutritional requirement for  
CC livestock and humans. The methods are also useful for the identification  
CC of pantothenate kinase modulators. Panto-compounds are produced at a  
CC significantly higher yield than prior art methods and can be produced  
CC independent of the need to feed precursors which decreases expense.  
XX  
SQ Sequence 258 AA;

Query Match 24.5%; Score 319; DB 22; Length 258;  
Best Local Similarity 28.1%; Pred. NO. 1.5e-28;  
Matches 76; Conservative 52; Mismatches 106; Indels 36; Gaps 4;  
QY 3 LVVDIGNTSTTLAIF-----TGDEEPSVESVPSALFADSSSTMREVFNG 45  
Db 3 LVIDVGNNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMI-----LRSLEFDH 49  
QY 46 MARKHGEPOAIAICSVVPSATAVGSALLESFVSPVLTICCKLRFPRLDYATPHFTGAD 105  
Db 50 SGLMFEQIDGIISSVPPIMFALERMCTKYFHEIQIVGPGMKTGLNICYDNPKREVGA 109  
QY 106 RLALCAWSRHLFSEKPKVIAVDIGTAITFDVLDTVGNVYRGGLIMPIDMMAGALHSRTAQL 165  
Db 110 RIVNAVAIAHLYG-NPLIVVDFGTATTTCYIDENKQYMGGAIAPIGTTISTEALYSRAAKL 168  
QY 166 PQVRIDRPESLLGRSTTECIKSGVFWGVVKQIGGLVDLRDVGESTVEIVTGGN 225  
Db 169 PRIETRPDNIIGKNTVSAMQSGILFGYVGVQVEGIVKRMKWQAKQD-----LKVIATGGL 223  
QY 226 SRIIVPEIGPVSVIDEAVLRGSDLLLRMN 255  
224 APLIANESDCIDIVDPFLTLKGLELIYERN 253

RESULT 4  
AAU91149  
ID AAU91149 standard; Protein; 258 AA.  
XX  
AC AAU91149;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Bacillus subtilis pantothenate kinase Coax.  
XX  
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX  
OS Bacillus subtilis.  
XX  
PN WO200216601-A2.  
XX  
PD 28-FEB-2002.  
XX  
PF 24-AUG-2001; 2001WO-US26531.  
XX  
PR 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.

XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
PA Yocum RR, Patterson TA;  
XX  
PI WPI; 2002-269358/31.  
XX N-PSDB; ABK54168.  
DR  
DR  
XX  
PT Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -  
XX  
PS Claim 10; Page 67-68; 128pp; English.  
XX  
CC The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.

XX  
SQ Sequence 258 AA;  
Query Match 24.5%; Score 319; DB 23; Length 258;  
Best Local Similarity 28.1%; Pred. NO. 1.5e-28;  
Matches 76; Conservative 52; Mismatches 106; Indels 36; Gaps 4;  
QY 3 LVVDIGNTSTTLAIF-----TGDEEPSVESVPSALFADSSSTMREVFNG 45  
Db 3 LVIDVGNNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMI-----LRSLEFDH 49  
QY 46 MARKHGEPOAIAICSVVPSATAVGSALLESFVSPVLTICCKLRFPRLDYATPHFTGAD 105  
Db 50 SGLMFEQIDGIISSVPPIMFALERMCTKYFHEIQIVGPGMKTGLNICYDNPKREVGA 109  
QY 106 RLALCAWSRHLFSEKPKVIAVDIGTAITFDVLDTVGNVYRGGLIMPIDMMAGALHSRTAQL 165  
Db 110 RIVNAVAIAHLYG-NPLIVVDFGTATTTCYIDENKQYMGGAIAPIGTTISTEALYSRAAKL 168  
QY 166 PQVRIDRPESLLGRSTTECIKSGVFWGVVKQIGGLVDLRDVGESTVEIVTGGN 225  
Db 169 PRIETRPDNIIGKNTVSAMQSGILFGYVGVQVEGIVKRMKWQAKQD-----LKVIATGGL 223  
QY 226 SRIIVPEIGPVSVIDEAVLRGSDLLLRMN 255  
Db 224 APLIANESDCIDIVDPFLTLKGLELIYERN 253

RESULT 5  
AAU91170  
ID AAU91170 standard; Protein; 262 AA.  
XX  
AC AAU91170;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Pantothenate kinase (Coax) #8.  
XX  
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX  
OS Bacillus anthracis.  
XX  
PN WO200216601-A2.  
XX  
PD 28-FEB-2002.  
XX  
PF 24-AUG-2001; 2001WO-US26531.  
XX





AC AAU911172;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Pantothenate kinase (Coax) #10.  
XX  
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX  
OS Bacillus stearothermophilus.  
XX  
PN WO200216601-A2.  
XX  
PD 28-FEB-2002.  
XX  
PF 24-AUG-2001; 2001WO-US26531.  
XX  
PR 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
PI Yocum RR, Patterson TA;  
XX WPI; 2002-269358/31.  
DR N-PSDB; ABK54193.  
XX  
PT Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein  
XX  
PS Claim 10; Page 101-102; 128pp; English.  
XX  
CC The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.  
XX  
SQ Sequence 258 AA;  
Query Match 23.4%; Score 305; DB 23; Length 258;  
Best Local Similarity 29.3%; Pred. No. 6.6e-27;  
Matches 76; Conservative 52; Mismatches 121; Indels 10; Gaps 4;  
QY 1 MRLVVDIGNSTTLAIFTGDE---EPSVESVPVSALFADSSSTM-REVFGNMARKHGEPAI 56  
Db 1 MIFVLDVGNNTNVLGVYDGDDELKHHWRIETSRKTEDEYGMKALLNHVGLQFSDIRGI 60  
QY 57 AICSVVPSATAVGSALLESLFSPVLTICCKLRFPRLDYATPHTFGADRLALCAWSRHL 116  
Db 61 IISVVPPIMFALERMCLKYFHKKPLIVGPIKTLGDIKIDNPREVGADRIVNAVAGIHL 120  
QY 117 FSEKPVIAVDIGTAITFDVLDTVGNRYRGLIMPIDMAGALHSRTAQLPQVRIDRPESL 176  
Db 121 YG-SPLIIVDFGTATTTCYINEHKQYMGGAIAPIGIMISTEALFARAALKPRIETARPDDI 179  
QY 177 LGRSTTECIKSGVFWGVKQIGGLVDAIRGDLVDFGDESTVEIVTGGNSRIIVPEIGPV 236  
Db 180 IGKNTVSAMQAGILYGVGVQVEGIVSRMKAK-----SKIPPKVIAATGGLAPLASESDII 234  
QY 237 SVIDELAVLRGSDLLLRMN 255  
Db 235 DVVDPFLLTLGLKLLYEKN 253  
RESULT 8  
AAU911154

ID AAU911154 standard; Protein; 255 AA.  
XX  
AC AAU911154;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Geobacter sulfurreducens pantothenate kinase Coax.  
XX  
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX  
OS Geobacter sulfurreducens.  
XX  
PN WO200216601-A2.  
XX  
PD 28-FEB-2002.  
XX  
PF 24-AUG-2001; 2001WO-US26531.  
XX  
PR 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
PI Yocum RR, Patterson TA;  
XX WPI; 2002-269358/31.  
XX  
PT Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein  
XX  
PS Claim 10; Page 72-73; 128pp; English.  
XX  
CC The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.  
XX  
SQ Sequence 255 AA;  
Query Match 23.1%; Score 301; DB 23; Length 255;  
Best Local Similarity 31.8%; Pred. No. 1.9e-26;  
Matches 85; Conservative 48; Mismatches 108; Indels 26; Gaps 8;  
QY 1 MRLVVDIGNSTTLAIFTGDEEPSVESVPVSALFADSSSTMREVFG-----NMARKHG---E 52  
Db 1 MLLVIDVGNTNIVLGIYDGERLVRDWRVST----DKARTTDEYGILINELFRLAGLGLDQ 56  
QY 53 PQAIAICSVPSATAVGSALLESLFSPVLTICCKLRFPRLDYATPHTFGADRL--ALC 110  
Db 57 IRAVLISSVVPPLTGVLERLSLGYFGMRPLVGVPGIKTGMPIQVDNPREVGADRIVNAVA 116  
QY 111 AWSRHLESEKPVIAVDIGTAITFDVLDTVGNRYRGLIMPIDMAGALHSRTAQLPQVRI 170  
Db 117 GYEKYRTS---LIIVDFGTATTFDYVNRKGEYCGGAIAPGLVISTEALFORASKLPRVDI 173  
QY 171 DRPESLLGRSTTECIKSGVFWGVKQIGGLVDAIRGDLVDFGDES--TVEIVTGGNSRI 228  
Db 174 IRPSAIARNTVNSMQAGIYYGVV---GLVDEI---VTRMKAESKDAPRVIAATGGLASL 226  
QY 229 IVPEIGPVSVIDEAVLRGSDLLLRMN 255  
Db 227 IAPESKTIEAVEEYLTLEGLRILYERN 253  
RESULT 9

AAU91151  
ID AAU91151 standard; Protein; 265 AA.  
XX  
AC AAU91151;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Streptomyces coelicolor pantothenate kinase Coax.

XX Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX Streptomyces coelicolor.  
OS  
XX  
PN WO200216601-A2.  
XX  
PD 28-FEB-2002.  
XX  
PF 24-AUG-2001; 2001WO-US26531.  
XX  
PR 24-AUG-2000; 2000US-227860P.  
XX 20-MAR-2001; 2001US-0813453.

(OMNI-) OMNIGENE BIOPRODUCTS INC.  
PA  
XX Yocum RR, Patterson TA;  
XX  
PI  
XX WPI; 2002-269358/31.

XX Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -  
XX  
PS Claim 10; Page 69-70; 128pp; English.  
XX  
CC The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.  
XX  
SQ Sequence 265 AA;

Query Match 22.6%; Score 295; DB 23; Length 265;  
Best Local Similarity 32.9%; Pred. No. 1e-25;  
Matches 94; Conservative 42; Mismatches 96; Indels 54; Gaps 13;  
QY 1 MRLVVDIGNTSTTLAIFTGDEEPSVE-----SVPSALFADSST-----MREVFGNMR 48  
Db 1 MLLTIDVGNTHTVLGLFDG--EDIVEHWRISTDSRRTADELAVLQGLMGHPLIGD--- 55  
QY 49 KHGEP-QAIAICSWVPSAT-----AVGSALLES--LFSVPVLTICCKLRFPFR 93  
Db 56 ELGDGIDGTAICATVPSVLHELREVTTRYGDVPAVLVEPGVKTGPILT----- 105  
QY 94 LDYATPHFGADRLALCAWSRHLSEKPVIAVDIGTAITFDVLDTVGNRYGGLIMPIDM 153  
Db 106 -DH--PKEVGADRIINAAVAVELYG-GPAIVVDFTATTTFDAVSARGEYIGGVIAPGIEI 161  
QY 154 MAGALHSRTAQLPQVRIDRPESLLGRSTTECIKSGVFWGVVKQIGGLVDAIRGDLVRDFG 213  
Db 162 SVEALGVKAQLRKIEVARPRSVIGKNTVEAMQSGIVGFAGQVDGVNRMARELADD-- 219  
QY 214 ESTVEVIVTGGNSRIIVPEIGPVSVIDE---LAVLRGSDLLLRNMN 256  
Db 220 PDDVTVIATGGAPMV---LGESSVIDEHEPWLTLMLGLRLVYERNV 262

RESULT 10  
AAU91152

ID AAU91152 standard; Protein; 272 AA.

XX  
AC AAU91152;  
XX  
DT 05-JUN-2002 (first entry)

XX Mycobacterium tuberculosis pantothenate kinase Coax.  
DE  
XX Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX  
OS Mycobacterium tuberculosis.  
XX  
PN WO200216601-A2.  
XX  
PD 28-FEB-2002.

XX 24-AUG-2001; 2001WO-US26531.  
XX  
PR 24-AUG-2000; 2000US-227860P.  
XX 20-MAR-2001; 2001US-0813453.

(OMNI-) OMNIGENE BIOPRODUCTS INC.  
PA  
XX Yocum RR, Patterson TA;  
XX  
PI  
XX WPI; 2002-269358/31.

XX Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -  
XX  
PS Claim 10; Page 70-71; 128pp; English.  
XX  
CC The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.  
XX  
SQ Sequence 272 AA;

Query Match 21.1%; Score 274.5; DB 23; Length 272;  
Best Local Similarity 29.3%; Pred. No. 2.6e-23;  
Matches 77; Conservative 52; Mismatches 125; Indels 9; Gaps 5;  
QY 1 MRLVVDIGNTSTTLAIFTGDEEPS--VES---VPSALFADSSTMRVFGNMARKHGEPQ 54  
Db 1 MLLAIDVRNTHTVVGLISGMKEHAKVVQQWRIRTESEVTADLAL-TIDGLIGEDSERLT 59  
QY 55 AIAICSVVPSATAVGSALLESLF-SVPVLTICCKLRFPRLDYATPHTFGADRLALCAWS 113  
Db 60 GTAALSTVPSVLHEVRIMLDQYWPSVPHVLIIEPGVRTGTPLLVDNPKEVGADRVNCLAA 119  
QY 114 RHLFSEKPVIAVDIGTAITFDVLDTVGNRYGGLIMPIDMMAGALHSRTAQLPQVRIDRP 173  
Db 120 YDRF-RKAAIVVDGSSICVDVVSAGFELGGATAPGVQVSSDAAAARSALRRVELARP 178  
QY 174 ESLLGRSTTECIKSGVFWGVVKQIGGLVDAIRGDLVRDFGESTVEVIVTGGNSRIIVPEI 233  
Db 179 RSVVGKNTVECMQAGAVFGFAGLVGDLVGRVIREDSVSGFVSDHDAIVATGHTAPLLLP 238  
QY 234 GPVSVIDEAVLRGSDLLLRNMN 256  
Db 239 HTVDHYDQHLTLQGLRLVFERNL 261

```
RESULT 11
AAU91163
  ID AAU91163 standard; Protein; 233 AA.
XX
AC
XX
AC
XX
AAU91163;
DT 05-JUN-2002 (first entry)
XX
DE Pantothenate kinase (Coax) #1.
XX
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Bacillus subtilis.
XX
PN WO200216601-A2.
XX
XX 28-FEB-2002.
  D
  X
  PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
DR N-PSDB; ABK54169.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein -
XX
PS Disclosure; Page 81-82; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
  X
  SQ Sequence 233 AA;
    Query Match 21.0%; Score 273.5; DB 23; Length 233;
    Best Local Similarity 28.5%; Pred. No. 2.7e-23;
    Matches 65; Conservative 42; Mismatches 90; Indels 31; Gaps 3;
QY 1 MRLVVDIGNTSITLAIF-----TGDEEPSVESVPSALEADSSSTMREVF 43
Db 1 MLLVIDVGNTNTVLGVYHDGKLEYHWRITSRHKTEDEFGMI-----LRSLF 47
QY 44 GNMARHGGEPOAIAICSVVPSATAYGSALLESLSFVPLTICCKLRFPFLDYATPHTFG 103
Db 48 DHSGLMFEQIDGIISVVPPIMFALERMCTKYFHIEPQIVGPMKTGLNIDPNKEVG 107
QY 104 ADRLALCAWSRHLFSEKPVIAVDIGTAITFDVLDTVGNRYRGGGLIMPIDMMAGALHSRTA 163
Db 108 ADRIVNAVAIIHLYG-NPLIVVDFTATTTCYIDENKQYMGGAIAPIGTTISTEALYSRAA 166
QY 164 QLPQVRIDRPESLGRSTTECIKSGVFWGVVVKQIGGLVDAIRGDLVRD 211
Db 167 KLPRIETRPDNIIGNKNTVSAMQSGILFGYGVGVGIVKRMKWAQKD 214
RESULT 12
```

```
AAG81225
ID AAG81225 standard; Protein; 272 AA.
XX
AC AAG81225;
XX
DT 04-SEP-2001 (first entry)
XX
DE Mycobacterium tuberculosis potential drug target protein SEQ ID 276.
XX
KW Drug target; growth; organism viability; characterisation.
XX
OS Mycobacterium tuberculosis.
XX
PN WO200135317-A1.
XX
PD 17-MAY-2001.
XX
PF 13-NOV-2000; 2000WO-US31152.
XX
PR 12-NOV-1999; 99US-0165086.
PR 12-NOV-1999; 99US-0165124.
PR 01-FEB-2000; 2000US-0179531.
XX
PA (REGC ) UNIV CALIFORNIA.
XX
PI Eisenberg D; Rotstein SH, Marcotte EM;
XX
DR WPI; 2001-329193/34.
DR N-PSDB; AAH52076.
XX
PT Identifying nucleotide or polypeptide sequence for use as drug target,
PT involves providing algorithm that analyzes a functional relationship
PT between nucleotide or polypeptide sequences, and comparing the
PT sequences -
XX
PS Disclosure; Page 188; 207pp; English.
XX
CC This invention relates to a method for identifying a nucleotide or
CC polypeptide sequence that may be a drug target, or essential for growth
CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
CC represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium
CC tuberculosis proteins which are potential drug targets. The DNA and
CC protein sequences are used to illustrate the method of the invention. The
CC method involves providing an unknown nucleotide or polypeptide sequences,
CC and comparing it to a number of sequences along with at least one
CC algorithm capable of analysing a functional relationship between
CC nucleotide and polypeptide sequences. The method is useful for
CC characterising the function of nucleic acids and polypeptides that may be
CC useful as a target for a drug or essential for the growth or viability of
CC an organism.
  X
  SQ Sequence 272 AA;
    Query Match 20.8%; Score 271.5; DB 22; Length 272;
    Best Local Similarity 29.1%; Pred. No. 5.9e-23;
    Matches 76; Conservative 52; Mismatches 124; Indels 9; Gaps 5;
QY 3 LVVDIGNTSITLAIFTGDEEPS--VES---VPSALFADSSSTMREVFGNMARKHGEPOAI 56
Db 3 LAIDVRNTHTVVGLLSGMKEHAKVVQOWRI RTESEVTADELAL-TIDGLIGEDSERLTGT 61
QY 57 AICSVVPSATAVGSALLESLSF-SVPVLTICCKLRFPFLDYATPHTFGADRLALCAWSRH 115
Db 62 AALSTVPSVLHEVRIMLDQWPSVPHVLIPEGVRTGIPLLVDNPKVCGADRIVNCLAAAYD 121
QY 116 LFSEKPVIAVDIGTAITFDVLDTVGNRYRGGGLIMPIDMMAGALHSRTAQLPQVRIDRPES 175
Db 122 RF-RKAAIIVDFGSSICVDVVSARKEFLGGAIAPGVQVSSDAAAARSAALRRVELARPRS 180
QY 176 LIGRSTTECIKSGVFWGVVVKQIGGLVDAIRGDLVRDVGESTVEIVTGSNSRIIVPEIGP 235
Db 181 VVGKNTVECMQAGAVFGFAGLVGDLVGRIRREDVSGFSVDHDAIVATGHTAPLLPELHT 240
```





```

Db          ::||| |||
            215 MKVIATGG 222

RESULT 15
AAU91173
ID      AAU91173 standard; Protein; 260 AA.
XX

```

RESULT 15  
AAU91173  
ID AAU91173 standard; Protein; 260 AA.  
XX

ID    AAU91173 standard; Protein; 260 AA.  
XX

**XX**

**XX**

05-JUN-2002	(first entry)
DT	
XX	
DE	Pantothenate kinase (CoaX) #11.

KW	Pantothenate kinase; CoaX; antimicrobial;
KW	pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX	
OS	Caulobacter crescentus.

WO200216601-A2.  
4X  
PD 28-FEB-2002.

PF 24-AUG-2001; 2001WO-US26531.  
XX  
PR 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.

PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX  
PI Yocum RR, patterson TA:

DR WPI; 2002-269358/31.  
DR N-PSDB; ABK54194.

Identifying potential antibiotic or antimicrobial agent, comprising contacting composition comprising pantothenate kinase (Coax) protein with test compound and identifying inhibitor of the Coax protein

PS Claim 10; Page 102-103; 128pp; English.

The invention describes assays for identifying a (potential) anti-infectious agent comprising contacting an assay composition comprising a pantothene kinase (Coax) protein with a test compound, and determining the activity of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulators can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Coax is the amino acid sequence of a pantothene kinase (Coax) protein described in the invention.

SQ	Sequence	260	AA;
----	----------	-----	-----

Query Match	20.0%;	Score 261;	DB 23;	Length 260;
Best Local Similarity	31.0%;	Pred. No. 9.3e-22;		
Matches 83;	Conservative	41;	Mismatches 118;	Indels 26;

	MRLVVDIGNSTTTLAIFTGDE-----EPSVESVPSPA----LFADSDSTMREVFGNMFM
Qy	: ::   :        :    :    ::    :
	MLLALTEQGNTNMTFAIHDGASWVAQWRSAATESSTRTADYEVVWLSQLLSMQOGLGFRA
Dd	

Qy	52	EPQALTAICSWPSATAVGSALLSFLSVPLTI--CCKLRFPRLDYATPHTFGAL
		:  :
Db	57	-IDAVIISWVPOSIFNLRLNRRRYFNVEPLVIGENAKIGIDVRIE--KPSEAGAD

QY 110 CAWSRHLFSEKPVIAVDIGTALTFTDVLDTVGNRYRGLIMPGLDMMAGALHSRTAQ

Db 114 AIGAMVY-PGPLVVIDSGTATFTDIVAADGAFEGGIIAPGINISMQALHEAAAKI

```

170 IDRP--ESLLGRSTTECIKSGVFWGVVKQIGGLVDAIRGLVDRDFGESTVEVIVTQ
      ||  ::  |  ::||||| :  |||  ::  ||  :  |||
173 IQRPAGNRIVGTDTVSAMQSGVFWGYISLIEGLVARIKAEK---GE-PMTVIATQ

```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:03:27 ; Search time 10.6025 Seconds  
(without alignments)  
2330.267 Million cell updates/sec

Title: US-09-813-453A-53  
Perfect score: 1303  
Sequence: 1 MRLVVDIGNTSTTLAIFTGD.....VIDELAVLRGSDLLLRNMMP 257

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	322	24.7	254	2 F83660	hypothetical prote
2	307	23.6	259	2 AF1102	conserved hypothet
3	307	23.6	259	2 AF1464	conserved hypothet
4	295	22.6	265	2 T36391	hypothetical prote
5	280.5	21.5	274	2 H86937	conserved hypothet
6	274.5	21.1	272	2 A70955	hypothetical prote
7	273.5	21.0	233	2 S66100	conserved hypothet
8	268	20.6	273	2 E97293	probable transcrip
9	261	20.0	261	2 B87489	transcription acti
10	253.5	19.5	262	2 E75516	conserved hypothet
11	228.5	17.5	246	2 D72320	conserved hypothet
12	225	17.3	229	2 E70465	hypothetical prote
13	218	16.7	276	2 AI2292	hypothetical prote
14	197.5	15.2	273	2 D71326	conserved hypothet
15	166.5	12.8	262	2 F70165	conserved hypothet
16	162.5	12.5	592	2 B81009	conserved hypothet
17	159.5	12.2	592	2 H82031	BirA protein/Bvg a
18	157	12.0	248	2 H83111	probable biotin-[a
19	146.5	11.2	257	2 S75559	hypothetical prote
20	144	11.1	242	2 A82637	hypothetical prote
21	136.5	10.5	267	2 I40327	conserved hypothet
22	108	8.3	209	2 H81382	baf protein - Bord
23	105.5	8.1	223	2 F64627	hypothetical prote
24	99.5	7.6	223	2 G71887	hypothetical prote
25	99	7.6	224	2 A99571	hypothetical prote
26	94.5	7.3	616	2 G98325	conserved hypothet
27	94.5	7.3	616	2 AG2957	probable c4-dicarb
28	93	7.1	347	2 E71826	hypothetical prote
29	92.5	7.1	747	2 C75545	rod shape-determin phosphoribosylform

30	91	7.0	317	2 G84414	homoserine dehydro
31	89.5	6.9	466	2 G87612	cytosol aminopepti
32	88.5	6.8	305	2 H89751	protein C33E10.2 [
33	88.5	6.8	781	2 A69312	DNA polymerase B1
34	88.5	6.8	786	2 JC1298	aculeacin-A acylas
35	88	6.8	433	2 D75632	probable hemolysin
36	87.5	6.7	249	2 G82808	toluene tolerance
37	87.5	6.7	974	2 T35045	bacteriophage phiC
38	87.5	6.7	974	2 T30204	pgl2 protein - Str
39	87.5	6.7	2658	2 A86216	protein T23G18.2 [
40	85.5	6.6	395	2 E90438	hypothetical prote
41	85.5	6.6	592	2 T04813	amidotransferase (
42	85.5	6.6	593	2 T48876	glutamine amidotra
43	85.5	6.6	835	1 W2BE51	gene 51 protein -
44	85	6.5	378	2 E64044	rod shape-determin
45	85	6.5	427	2 C72618	probable 3-phospho

ALIGNMENTS

RESULT 1  
F83660  
hypothetical protein BH0086 [imported] - Bacillus halodurans (strain C-125)  
C;Species: Bacillus halodurans  
C;Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C;Accession: F83660  
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.;  
Nucleic Acids Res. 28, 4317-4331, 2000  
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans  
A;Reference number: A83650; MUID:20512582; PMID:11058132  
A;Accession: F83660  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-254 <STO>  
A;Cross-references: GB:AP001507; GB:BA000004; NID:g10172612; PIDN:BAB03805.1; GSPDB  
A;Experimental source: strain C-125  
C;Genetics:  
A;Gene: BH0086  
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match	24.7%	Score 322;	DB 2;	Length 254;
Best Local Similarity	28.9%	Pred. No. 1.8e-21;		
Matches	79;	Conservative	54;	Mismatches 104; Indels 36; Gaps 4;
QY	1	MRLVVDIGNTSTTLAIF-----TGDEEPSVESVPSALFADSSSTMREV	43	
Db	1	MILVIDGNTNTVLGVYQDETLVHHWRLATSRQKTEDEY-----AMTVRS	47	
QY	44	GNNARKHGEPOAIAICSVVPSATAVGSALLESLSFVPLTICCKLRFPRDLDYATP	103	
Db	48	DHAGLQFQDIDGIVISSVPPPMFSLQMKCKYFHVTPMIIGPGIKTGLNINIKYDNPKE	107	
QY	104	ADRLALCAWSRHLFSEKFPVIAVDIGTAITFDVLDTVGNVYRGGLIMPIDMAGALHS	163	
Db	108	ADRIVNAVAAIELYG-YPAIVVDFGTATTYCLINEKKQYAGGVIAAPGIMISTEALYH	166	
QY	164	QLPQVRIDRPESLLGRSTTECIKSGVFWGVVKQIGGLVDAIRGDLVRDFGESTVEIV	223	
Db	167	KLPRIETAKPKQVVGNTFIDSMQSGIFYGYVSQVDGVVKRMAQ-----AESEPK	221	
QY	224	GNSRIIVPEIGPVSVIDEAVLRGSDLLLRNMN	256	
Db	222	GLAKLIGTESETIDVIDSFLTKGLQLIYKKNV	254	

RESULT 2  
AF1102  
conserved hypothetical protein lmo0221 [imported] - Listeria monocytogenes (strain F  
C;Species: Listeria monocytogenes  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C;Accession: AF1102  
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Blo

.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A;Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
A;Title: Comparative genomics of Listeria species.  
A;Reference number: AB1077; MUID:21537279; PMID:11679669  
A;Accession: AF1102  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-259 <GLA>  
A;Cross-references: GB:NC\_003210; PIDN:CAD00748.1; PID:gl6409586; GSPDB:GN00177  
A;Experimental source: strain EGD-e  
C;Genetics:  
A;Gene: lmo0221  
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 23.6%; Score 307; DB 2; Length 259;  
Best Local Similarity 30.4%; Pred. No. 4.2e-20;  
Matches 83; Conservative 49; Mismatches 107; Indels 34; Gaps 7;  
/ 1 MRLVVDIGNTSTTLAIF-----TGDEEPSVESVPSALFADSSSTMREVFGNMARK 49  
1 MILVIDGNTNCTGVYKEQKLLKHWRMTTDRHRTSDEL-----GMTVLNFFSYANLT 53  
Db  
QY 50 HGEPOAIAICSVVPSATAVGSALLESLFSPVLTICCKLRFPPRLDYATPHTFGADRL-- 107  
54 PSDIQIGIISSVVPPIHMETMVCYRNFIRPLIVGPIKTLNKLKVDNPREIGSDRIVN 113  
Db  
QY 108 ALCAWSRHLFSEK---PVIADVIGTAITFDVLTGVNRYGGGLIMPIDMMAGALHSRTAQ 164  
114 AVAA-----SEYGTPIVVDFGTATTCFYIDESGVYQGGAIAPGIMISTEALYNRAAK 167  
Db  
QY 165 LPQVRIDRPESLLGRSTTECIKSGVFWGVVKQIGLVDAIRGDLVRDFGEGSTVEIVTGG 224  
168 LPRVDIAESSQIIGKSTVSMQAGIFYGFVQCEGII----AEMKKQSNASPV-VVATGG 222  
Db  
QY 225 NSRIIVPEIGPVSVIDEAVLRGSDLLLRNMMP 257  
223 LARMITEKSSAVDILDPLTLKGLLELYRRNKP 255  
Db

RESULT 3  
AF1464  
conserved hypothetical protein lin0253 [imported] - Listeria innocua (strain Clip11262)  
C;Species: Listeria innocua  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C;Accession: AF1464  
.; Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A;Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
A;Title: Comparative genomics of Listeria species.  
A;Reference number: AB1077; MUID:21537279; PMID:11679669  
A;Accession: AF1464  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-259 <GLA>  
A;Cross-references: GB:AL592022; PIDN:CAC95486.1; PID:gl6412682; GSPDB:GN00178  
A;Experimental source: strain Clip11262  
C;Genetics:  
A;Gene: lin0253  
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c  
Query Match 23.6%; Score 307; DB 2; Length 259;  
Best Local Similarity 30.0%; Pred. No. 4.2e-20;  
Matches 82; Conservative 50; Mismatches 107; Indels 34; Gaps 7;  
QY 1 MRLVVDIGNTSTTLAIF-----TGDEEPSVESVPSALFADSSSTMREVFGNMARK 49  
1 MILVIDGNTNCTGVYKEQKLLRHWRMTTDRHRTSDEL-----GMTVLNFFSYANLT 53  
Db

QY 50 HGEPOAIAICSVVPSATAVGSALLESLFSPVLTICCKLRFPPRLDYATPHTFGADRL-- 107  
Db 54 PSDIQIGIISSVVPPIHMETMVCYRNFIRPLIVGPIKTLNKLKVDNPREIGSDRIVN 113  
QY 108 ALCAWSRHLFSEK---PVIADVIGTAITFDVLTGVNRYGGGLIMPIDMMAGALHSRTAQ 164  
Db 114 AVAA-----SEYGTPIVVDFGTATTCFYIDEAGVYQGGAIAPGIMISTEALYNRAAK 167  
QY 165 LPQVRIDRPESLLGRSTTECIKSGVFWGVVKQIGLVDAIRGDLVRDFGEGSTVEIVTGG 224  
Db 168 LPRVDIAESSQIIGKSTVSMQAGIFYGFVQCEGII----AEMKKQSNTPSPV-VVATGG 222  
QY 225 NSRIIVPEIGPVSVIDEAVLRGSDLLLRNMMP 257  
Db 223 LARMITEKSSAVDILDPLTLKGLLELYRRNKP 255

RESULT 4  
T36391  
hypothetical protein SCE94.31c - Streptomyces coelicolor  
C;Species: Streptomyces coelicolor  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jun-2000  
C;Accession: T36391  
R;Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.  
submitted to the EMBL Data Library, April 1999  
A;Reference number: Z21573  
A;Accession: T36391  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-265 <OLI>  
A;Cross-references: EMBL:AL049628; PIDN:CAB40880.1; GSPDB:GN00070; SCOEDB:SCE94.31c  
A;Experimental source: strain A3(2)  
C;Genetics:  
A;Gene: SCOEDB:SCE94.31c  
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 22.6%; Score 295; DB 2; Length 265;  
Best Local Similarity 32.9%; Pred. No. 5.3e-19;  
Matches 94; Conservative 42; Mismatches 96; Indels 54; Gaps 13;  
QY 1 MRLVVDIGNTSTTLAIFTGDEEPSVE---SVPSALFADSSST-----MREVFGNMAR 48  
Db 1 MLITIDVGNTHTVLGLFDG--EDIVEHWRISTDSRRRTADELAVLQGLMGHPLIGD--- 55  
QY 49 KHGEP-QAIAICSVVPSAT-----AVGSALLES--LFSVPVLTICCKLRFPPFR 93  
Db 56 ELGDGIDGIAICATVPVSVLHELREVTTRRYGDPVAVLVEPGVKTGPILT----- 105  
QY 94 LDYATPHTFGADRLALCAWSRHLFSEKPKVIAVDIGTAITFDVLTGVNRYGGGLIMPIDM 153  
Db 106 -DH--PKEVGADRIINAVAAVELYG-GPAIVVDFGTATTTFDAVSARGEYIGGVIAPGIEI 161  
QY 154 MAGALHSRTAQLPQVRIDRPESLGRSTTECIKSGVFWGVVKQIGLVDAIRGDLVRDFG 213  
Db 162 SVEALGVKGAQLRKIEVARPRSVIGKNTVEAMQSGIVYGFAGQVDGVVNRMARELADD-- 219  
QY 214 ESTVEIVTGGNSRIIVPEIGPVSVIDE---LAVLRGSDLLLRNM 256  
Db 220 PDDVTVIATGGLAPMV---LGESSVIDEHEPWLTMLGLRLVYERNV 262

RESULT 5  
H86937  
conserved hypothetical protein ML0232 [imported] - Mycobacterium leprae  
C;Species: Mycobacterium leprae  
C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
C;Accession: H86937  
R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.  
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holr  
eam, M.A.; Rutherford, K.M.  
Nature 409, 1007-1011, 2001  
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.





RESULT 8  
E97293  
probable transcription regulator, homolog of Bvg accessory factor [imported] - Clostridi  
C;Species: Clostridium acetobutylicum  
C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001  
C;Accession: E97293  
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo  
A;Reference number: A96900; MUID:21359325; PMID:21359325  
A;Accession: E97293  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-273 <KUR>  
A;Cross-references: GB:AE001437; PIDN:AAK81136.1; PID:g15026270; GSPDB:GN00168  
A;Experimental source: Clostridium acetobutylicum ATCC824  
C;Genetics:  
Gene: CAC3200  
Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c  
Query Match 20.6%; Score 268; DB 2; Length 273;  
Best Local Similarity 28.4%; Pred. No. 1.5e-16;  
Matches 74; Conservative 59; Mismatches 114; Indels 14; Gaps 6;  
QY 3 LVVDIGNTTTLAIF-----TGDEEPSVESVPSALFADSSSTMREVFGNMARKHGE--Q 54  
Db 14 LVLDVGNNTNIVLGIYNDTKLTAEWRLSTDVLS---ADEYGI-QVMNLFQDKLDPTLVE 69  
QY 55 AIAICSVVPSAFVGSALLESLFVSVPLTICCKLRFPFRLDYATPHTFGADRLALCAWSR 114  
Db 70 GVLISSVVPNIMYSLEHMIRKYFKINPLVVGPGTKTGINIKYDNPKEVGADRI-VNAVA 128  
QY 115 HLFSEKPVIAVDIGTAITFDVLDTVGNRYGGLIMPIDMAGALHSRTAQLPQVRIDRPE 174  
Db 129 HEIYKRSLLIIDFGTATTFCAVRENGDYLGAICPGIKVSSEALFEKAALPRVELIKPA 188  
QY 175 SLGRSTTECIKSGVFWGVVKQIGGLVDAIRGDLVRDFGEGSTVEIVTGGNSRIIVPEIG 234  
Db 189 YAICKNTISSQSGIVGYIGQVRYIVERMKEEL-QEEGEKEPLVAVTGGLAKLISEAK 247  
QY 235 PVSVIDELAVLRGSDLLLRMN 255  
Db 248 NVDVINPFLTLEGLRIIYEKN 268  
RESULT 9  
7489  
transcription activator, probable Baf family [imported] - Caulobacter crescentus  
C;Species: Caulobacter crescentus  
C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
C;Accession: B87489  
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A;Title: Complete Genome Sequence of Caulobacter crescentus.  
A;Reference number: AB7249; MUID:21173698; PMID:11259647  
A;Accession: B87489  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-261 <STO>  
A;Cross-references: GB:AE005673; NID:g13423392; PIDN:AAK23910.1; GSPDB:GN00148  
C;Genetics:  
A;Gene: CC1935  
Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c  
Query Match 20.0%; Score 261; DB 2; Length 261;  
Best Local Similarity 31.0%; Pred. No. 6e-16;  
Matches 83; Conservative 41; Mismatches 118; Indels 26; Gaps 9;

QY 1 MRLVVDIGNTTTLAIFTGDE-----EPSVESVPSA-----LFADSSSTMREVFGNMARKHG 51  
Db 2 MLLATEQGNNTMTFAIHGASVWQWRSATSTRTADEYVWVLSQLSMQGLGFRA----- 57  
QY 52 EPQAIAICSVVPSATAVGSALLESLFVSVPLTI--CCKLRFPFRLDYATPHTFGADRLAL 109  
Db 58 -IDAVLISSVVPQSIENLRLSRFVNEPLVIGENAKLIGDIVRIE--KPSEAGADRLVN 114  
QY 110 CAWSRHLFSEKPVIAVDIGTAITFDVLDTVGNRYGGLIMPIDMAGALHSRTAQLPQVR 169  
Db 115 AIGAAMVY-PGPLVVIDSGTATFDIVAADGAFEGGIIAPGINLSMOALHEAAKLPRIA 173  
QY 170 IDRP--ESLLGRSTTECIKSGVFWGVVKQIGGLVDAIRGDLVRDFGEGSTVEIVTGGNSR 227  
Db 174 IQRPAGNRIIVGTDVTSAMQSGVFWGYISLIEGLVARIKAER----GE-PMTVIATGGVAS 228  
QY 228 IIVPEIGPVSVIDEAVLRGSDLLLRMN 255  
Db 229 LFEGATDSIDHFDSDLTIRGLLEIYRN 256  
RESULT 10  
E75516  
conserved hypothetical protein - Deinococcus radiodurans (strain R1)  
C;Species: Deinococcus radiodurans  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 20-Jun-2000  
C;Accession: E75516  
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.  
; M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A;Reference number: A75250; MUID:20036896; PMID:10567266  
A;Accession: E75516  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-262 <WHI>  
A;Cross-references: GB:AE001905; GB:AE000513; NID:g6458144; PIDN:AAF10040.1; PID:g64  
A;Experimental source: strain R1  
C;Genetics:  
A;Gene: DR0461  
A;Map position: 1  
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c  
Query Match 19.5%; Score 253.5; DB 2; Length 262;  
Best Local Similarity 27.8%; Pred. No. 2.8e-15;  
Matches 74; Conservative 46; Mismatches 103; Indels 43; Gaps 7;  
QY 3 LVVDIGNTTTLAIFTGDEEPSVESVPSALFADSS-----TMREVFGN--MARKH 50  
Db 7 LAVDIGNTTTVLGL-----ADASGALHTWRTNREMLPDDDLALQLH 49  
QY 51 G-----EPQAIAICSVVPSATAVGSALLESLFVSVPLTICCKLRFPFRLDYATPHTF 102  
Db 50 GLFTLAGAPIPRAAVLSSVAPPVGENYALAKRHFMDAFVSAENLPDVTVELDTPGSV 109  
QY 103 GADRLA-LCAWSRHLFSEKPVIAVDIGTAITFDVLDTVGNRYGGLIMPIDMAGALHSR 161  
Db 110 GADRLCNLFGAEKYLGGLDYAVVDFGTSTNFDVVGRGRRLGGILATGAQVVSADALFAR 169  
QY 162 TAQLPQVRIDRPESLLGRSTTECIKSGVFWGVVKQIGGLVDAIRGDLVRDFGEGSTVEIV 221  
Db 170 AAKLPRIITLQAPETAIGKNTVHALQSLGVFGYAEVMDGLLRIRRAELP---GEAV--AVA 224  
QY 222 TGGNSRIIVPEIGPVSVIDEAVLRG 247  
Db 225 TGGFSRTVQGICQEIYDYEDELTLRG 250  
RESULT 11  
D72320  
conserved hypothetical protein - Thermotoga maritima (strain MSB8)  
C;Species: Thermotoga maritima







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:51:40 ; Search time 5.16979 Seconds  
(without alignments)  
2061.866 Million cell updates/sec

Title: US-09-813-453A-53

Perfect score: 1303

Sequence: 1 MRLVVDIGNTSTTLAIFTGD.....VIDELAVIRGSDLLLRNMMP 257

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB Length	ID	Description
1	273.5	21.0	233	1	YACB_BACSU	P37564 bacillus su
2	136.5	10.5	267	1	BAF_BORPE	Q45338 bordetella
3	92.5	7.1	747	1	PURL_DEIRA	Q9rxt4 deinococcus
4	91.5	7.0	554	1	SECD_RHOCA	O33517 rhodobacter
5	88.5	6.8	781	1	DPOL_ARCFU	O29753 archaeoglob
6	88.5	6.8	786	1	AAC_ACTUT	P29958 actinoplane
7	85.5	6.6	592	1	HIS5_ARATH	Q9sz30 arabidopsis
8	85.5	6.6	835	1	OBP_VZVD	P09299 varicella-z
9	85	6.5	351	1	MREB_HAEIN	P44474 haemophilus
10	82.5	6.3	810	1	SYFB_SYNY3	P74296 synechocyst
11	82	6.3	351	1	MREB_PASMU	Q916a3 pasteurella
12	81	6.2	260	1	HGDC_ACIFE	P11568 acidaminoco
13	81	6.2	1106	1	ACLY_CAEEL	P53585 caenorhabdi
14	80.5	6.2	384	1	NIFS_RHOCA	Q07177 rhodobacter
15	80	6.1	331	1	PGTB_RAT	Q08603 rattus norv
16	80	6.1	488	1	BTB1_MOUSE	P58544 mus musculu
17	79.5	6.1	524	1	AMPA_AERPE	Q9Y935 aeropyrum p
18	79	6.1	337	1	MREB_BACSU	Q01465 bacillus su
19	79	6.1	509	1	YMCB_BACSU	O31778 bacillus su
20	79	6.1	1693	1	POLN_HEVBU	P29324 hepatitis e
21	79	6.1	1693	1	POLN_HEVBY	Q04610 hepatitis e
22	79	6.1	2021	1	OMPA_RICCN	Q52657 rickettsia
23	78.5	6.0	296	1	NPL_ECOLI	P06995 escherichia
24	78.5	6.0	533	1	PEPC_ASPNG	P33295 aspergillus
25	78	6.0	373	1	PRXC_CALFU	P04963 caldariomyc
26	78	6.0	444	1	YOJA_BACSU	O31862 bacillus su
27	78	6.0	495	1	AMPA_PSEAE	O68822 pseudomonas
28	78	6.0	575	1	SMF1_YEAST	P38925 saccharomyc
29	78	6.0	634	1	DNAK_HAEIN	P43736 haemophilus
30	78	6.0	760	1	METE_MYCLE	O05564 mycobacteri
31	77.5	5.9	528	1	UDPE_NPVM	Q83140 mamestra br
32	77.5	5.9	1403	1	VG22_HSV11	Q00105 ictaluriid h
33	77.5	5.9	4367	1	DYHC_NEUCR	P45443 neurospora

34	77	5.9	288	1	BLC6_VIBCH	P81781 vibrio chol
35	77	5.9	338	1	PSC1_MYCTU	P95303 mycobacteri
36	77	5.9	428	1	BCH2_RHOCA	P26171 rhodobacter
37	77	5.9	799	1	YDDU_ECOLI	P76129 escherichia
38	77	5.9	1037	1	YOJ8_YEAST	Q12496 saccharomyc
39	76.5	5.9	278	1	CYST_SYNP7	P27367 synechococc
40	76.5	5.9	342	1	HEMZ_RICPR	Q9zc84 rickettsia
41	76.5	5.9	463	1	VL2_HPVA5	P36761 human papil
42	76.5	5.9	916	1	PMAL_AJECA	Q07421 ajellomyces
43	76	5.8	360	1	BUK_ENTFA	Q9rps7 enterococcu
44	76	5.8	591	1	KPY2_SYNY3	P73534 synechocyst
45	76	5.8	1100	1	TRAA_AGR75	Q44349 agrobacteri

ALIGNMENTS

RESULT 1  
YACB\_BACSU  
ID YACB\_BACSU STANDARD; PRT; 233 AA.  
AC P37564;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein yacB.  
GN YACB.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=96051385; PubMed=7584024;  
RA Ogasawara N., Nakai S., Yoshikawa H.;  
RT "Systematic sequencing of the 180 kilobase region of the Bacillus  
RT subtilis chromosome containing the replication origin.";  
RL DNA Res. 1:1-14(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
RA Entlan K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
RA Guiseppe G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,  
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
RA Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,  
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
RA Sorokia A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
RT subtilis.";  
RL Nature 390:249-256(1997).  
RL -!- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; D26185; BAA05305.1; -  
EMBL; 299104; CAB11846.1; -  
Subtilist; BG10133; yacB.  
InterPro; IPR004619; Baf.  
Pfam; PF03309; Bvg\_acc\_factor; 1.  
TIGRFAMs; TIGR00671; baf; 1.  
Hypothetical protein; Complete proteome.  
SEQUENCE 233 AA; 26217 MW; AAE96E732C15DF44 CRC64;  
  
Query Match 21.0%; Score 273.5; DB 1; Length 233;  
Best Local Similarity 28.5%; Pred. No. 4.7e-17;  
Matches 65; Conservative 42; Mismatches 90; Indels 31; Gaps 3;  
  
1 MRLVVDIGNTSTTIAIF-----TGDEEPSVESVPSALFADSTSMREVF 43  
1 MLLVIDVGNTNTVLGVYHDKLEYHWRIETSRHKTDEFGMI-----LRSLF 47  
44 GNMARHGEPQAIACISVVPSATAVGSALLESFSPVLTICCKLRFPRLDYATPHTFG 103  
48 DHSGLMFEQIDGIIISVVPPIMFALERMCTKYFHIEPQIVGPMKGTGLNKNYDNPEVG 107  
104 ADRLALCAWSRHLFSEKPVIAVDIGTAITFDVLTGVNVRGGLIMPIDMMAGALHSRTA 163  
108 ADRIVNAVAIAHLYG-NPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAA 166  
164 QLPQVRIDRPESLLGRSTTECIKSGVFWGVVKQIGGLVDAIRGDLVRD 211  
167 KLPRIETRPDNIIGKNIVSAMQSGILFGYVGVQVEGIVKRMKQAKQD 214  
  
RESULT 2  
BAF\_BORPE STANDARD; PRT; 267 AA.  
ID BAF\_BORPE Q45338; Q45373;  
AC Q45338; Q45373;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Bvg accessory factor.  
GN BAF.  
OS Bordetella pertussis.  
Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;  
Bordetella.  
NCBI\_TaxID=520;  
[1]  
SEQUENCE FROM N.A.  
STRAIN-BP504;  
MEDLINE-95325323; PubMed-7601846;  
Deshazer D., Wood G.E., Friedman R.L.;  
"Identification of a Bordetella pertussis regulatory factor required  
for transcription of the pertussis toxin operon in Escherichia  
coli.";  
J. Bacteriol. 177:3801-3807(1995).  
[2]  
SEQUENCE OF 1-38 FROM N.A.  
STRAIN-BP504;  
Wood G.E., Friedman R.L.;  
"Identification of a bira homolog in Bordetella pertussis.";  
Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
[3]  
SEQUENCE OF 239-267 FROM N.A.  
STRAIN-BP536;  
MEDLINE-96419162; PubMed-8821935;  
Allen A.G., Maskell D.J.;  
"The identification, cloning and mutagenesis of a genetic locus

required for lipopolysaccharide biosynthesis in Bordetella  
pertussis.";  
Mol. Microbiol. 19:37-52(1996).  
-!- FUNCTION: ACTIVATES TRANSCRIPTION OF THE PERTUSSIS TOXIN OPERON IN  
A BVGAS-DEPENDENT MANNER. MAY INTERACT WITH THE ALPHA SUBUNIT OF  
RNA POLYMERASE.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; U12020; AAA75361.1; -  
EMBL; AF016461; AAC68834.1; -  
EMBL; X90711; CAA62242.1; -  
InterPro; IPR004619; Baf.  
Pfam; PF03309; Bvg\_acc\_factor; 1.  
Transcription regulation; Activator.  
SEQUENCE 267 AA; 27845 MW; 336AG15F67B57901 CRC64;  
  
Query Match 10.5%; Score 136.5; DB 1; Length 267;  
Best Local Similarity 26.2%; Pred. No. 6.8e-05;  
Matches 62; Conservative 31; Mismatches 113; Indels 31; Gaps 9;  
  
1 MRLVVDIGNTSTTIAIFTGDEEPSVESVPSALFADSTSMREVFGNMARKHGEQP----- 54  
1 MIILDSGNSRLKVGWFDPD-APQAREPAPVAFDNLDLALGRWLATLPRRPQALGVN 59  
55 --ATAICSVVPSATAVGSALLESFSPVLTICCKLRFPRLDYATPHTFGADRLA--LC 110  
60 VAGLARGEAIAATLRAGGCDIRWLRAP-----LAMGLRNGYRNPDLGADRWACMVG 112  
111 AWSRHLFSEKPVIAVDIGTAITFDVLTGVNVRGGLIMPIDMMAGALHSRTAQLP---Q 167  
113 VLARQPSVHPPLLVASFSGTATTTLDTIGPDNVFPFPGGLILPGPAMMRGALAYGTAHLPLADG 172  
168 VRIDRPESLLGRSTTECIKSGVFWGVVKQIGGLVDAIRGDLVRDFGSESTVEVITGG 224  
173 LVADYP-----IDTHQAIASGI---AAAQAGAIVRQWLAGRQR-YGQAP-EIYVAGG 219  
  
RESULT 3  
PURL\_DEIRA STANDARD; PRT; 747 AA.  
ID PURL\_DEIRA Q9RXT4;  
AC Q9RXT4;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Phosphoribosylformylglycinamide synthase II (EC 6.3.5.3) (FGAM  
synthase II).  
GN PURL OR DR0222.  
OS Deinococcus radiodurans.  
Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;  
OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;  
OC Deinococcales; Deinococcus.  
OX NCBI\_TaxID=1299;  
RN [1]  
SEQUENCE FROM N.A.  
STRAIN-RL;  
RC STRAIN-RL;  
RX MEDLINE-20036896; PubMed-10567266;  
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,  
Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
Fraser C.M.;  
"Genome sequence of the radioresistant bacterium Deinococcus  
radiodurans R1.";  
Science 286:1571-1577(1999).  
-!- CATALYTIC ACTIVITY: ATP + N(2)-formyl-N(1)-(5-phospho-D-

```
CC ribosyl)glycinamide + L-glutamine + H(2)O = ADP + phosphate + 2-
CC (formamido)-N(1)-(5-phospho-D-ribose)acetamidine + L-glutamate.
CC -!- PATHWAY: De novo purine biosynthesis; fourth step.
CC -!- SUBUNIT: HETERODIMER OF TWO SUBUNITS, PURQ AND PURL.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE FGAMS FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE001884; AAF09808.1; -
CC TIGR; DR0222; -
CC InterPro; IPR000728; AIRS_related.
CC Pfam; PF00586; AIRS; 2.
CC Pfam; PF02769; AIRS_C; 2.
CC Purine biosynthesis; Ligase; ATP-binding; Complete proteome.
CC NP_BIND 103 114 ATP (POTENTIAL).
CC SEQUENCE 747 AA; 78366 MW; CB169EEC6E063724 CRC64;
SQ
Query Match 7.1%; Score 92.5; DB 1; Length 747;
Best Local Similarity 22.1%; Pred. No. 1.7;
Matches 64; Conservative 38; Mismatches 102; Indels 85; Gaps 14;
QY 4 VVDIGNTSTLAFTGDEEPS-VESVPSALFADSSSTMREVFGNMR-----KHGEPO 54
Db |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
76 VVDIGDGNVAFKMFESHNPSSAVEPVQGAATGVGGILRDIFAMGARPFVLDLSLRFNGPD 135
QY 55 A-----IAICSVVPSATAVGSALLESFSPVLTICCKLRFPFRDLYATPHTFGADRLALC 110
Db : || : || : || : || : || : || : || : || : || : || : || : || : ||
136 SPRTFLVNGVVDGIAHYGNAI-----GVP--TVGGEVTF----- 168
QY 111 AWSRHLFSEKPIAV-----DIGTAITFDVLDTVGNVRGGLIMPIDM 153
Db : || : || : || : || : || : || : || : || : || : || : || : || : ||
169 ---HPSYQENPLNVNVMALGLLRHEDLATCPMGEVGNQIVYVGSKTGRDGLGGAVFSSADL 225
QY 154 MAGALHSRTAQLPQVRIDRP--ESLLGRSTTECIKSGVFWGVKQIG--GLVD-----AI 204
Db | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
226 SAAS---QADRPVQVGDPPFMEKLLLEATLEAIQAGLVAG-VQDMGAAGLVSSCTEMAY 280
QY 205 RGDG-----VRDFGESTVEVITGNSR-IIVPEIGPVSVIDEI 242
Db | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
281 RASLGITMDLDKVPTRREGMVPMECLSESQERMILVPVPGKEQALHDL 329
```

RESULT 4

```
SECD_RHOCA STANDARD; PRT; 554 AA.
ID SECD_RHOCA
AC O33517;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Protein-export membrane protein secD.
GN SECD.
OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
OX NCBI_TaxID=1061;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SB1003 / St Louis;
RX MEDLINE=97318920; PubMed=9175857;
RA Goldman B.S., Beckman D.L., Bali A., Monika E.M., Gabbert K.K.,
RA Kranz R.G.;
RT "Molecular and immunological analysis of an ABC transporter complex
RT required for cytochrome c biogenesis.";
RL J. Mol. Biol. 268:724-738(1997).
CC -!- FUNCTION: INVOLVED IN PROTEIN EXPORT (BY SIMILARITY).
CC -!- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS
```

```
CC WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECY
CC (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECD FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U69979; AAB62801.1; -
CC InterPro; IPR003335; SecD_SecF.
CC Pfam; PF02355; SecD_SecF; 1.
CC TIGRFAMS; TIGR00916; 2A0604s01; 1.
CC TIGRFAMS; TIGR01129; secD; 1.
CC Protein transport; Translocation; Transmembrane; Membrane.
CC TRANSMEM 10 30 POTENTIAL.
CC TRANSMEM 392 412 POTENTIAL.
CC TRANSMEM 435 455 POTENTIAL.
CC TRANSMEM 491 511 POTENTIAL.
CC TRANSMEM 516 536 POTENTIAL.
CC SEQUENCE 554 AA; 58943 MW; DF2CBEEBA9F69EDF CRC64;
SQ
Query Match 7.0%; Score 91.5; DB 1; Length 554;
Best Local Similarity 24.5%; Pred. No. 1.5;
Matches 67; Conservative 39; Mismatches 111; Indels 57; Gaps 16;
QY 16 IFTGDE----EPSVE--SVPSALFADSSSTMREVFGNMARKH-CEPQAIACSVVPSATAV 68
Db : |||: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
278 VVTGDDLTDARPTTDDNGAPAVSFRFNVSGARAFGDYTAGHIGEPFAIVLDGKVISAPTI 337
QY 69 -----GSALLESFSPVLT-ICCKLR---FPRLDYATPHT-----FGADR----- 106
Db || : || : || : || : || : || : || : || : || : || : || : || : ||
338 QAHIAGSGGIITGRFSIEETDLALLRAGALPAGMTFLEERTIGPELGADSVKAGMVAS 397
QY 107 ----LALCAW---SRHLSEKPIAVDIGTAITFDVLDTVGNVRGGLIMPIDMAGAL- 158
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
398 VIGFVAVVAYMIASYLEGFFSSVALFINIAFIFAVMGAIG---GTMTLPGI---AGIVL 451
QY 159 ---HSRTAQLPQVRIDRPESLLGRSTTECIKSGVFWGVKQIGGLVDAIRGLVDRDFGES 215
Db | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
452 TIGTSVDANVLIYERMREIRSGKSPVRAIE---LGFDKAMSAIIDA----NVTSLSS 503
QY 216 TVEVITGNSR--IIVPEIG-PVSVIDEI 246
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
504 AILFVLGAGPVRGFAVTTMIGTAASIFTAIWVR 537
```

RESULT 5

```
DPOL_ARCFU STANDARD; PRT; 781 AA.
ID DPOL_ARCFU
AC O29753;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA polymerase (EC 2.7.7.7).
GN POL OR POLB OR AF0497.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
```









CC BHV-1 53, AND VZV 51.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----

DR EMBL; X04370; CAA27933.1; -  
DR PIR; G27344; WZBES1.  
DR InterPro; IPR003450; Herpes\_ori\_bp.  
DR Pfam; PF02399; Herpes\_ori\_bp; 1.  
KW DNA replication; DNA-binding; ATP-binding.  
FT NP\_BIND 67 74  
SQ SEQUENCE 835 AA; 94374 MW; A71F5877ACF386FB CRC64;

Query Match 6.6%; Score 85.5; DB 1; Length 835;  
Best Local Similarity 22.1%; Pred. No. 8.2;  
Matches 55; Conservative 30; Mismatches 93; Indels 71; Gaps 9;

QY 49 KHGEPQAIACSV-VPSATAVGSALLESL-----FSVPVLTICCKLRFPPRL----- 94  
Db 53 KPGMSQTRPVTVRAPMGSGKTTALLEWLQALKADISLVVSCRRSFTQTLLIQRFNDAG 112  
QY 95 -----DYATPHTF--GADRLALCAWSRHLSEKPVIAVDIGTAITFDVLTGVNRYGGGLI 147  
Db 113 LSGFVTYLTSETYINGFKRLIVQLVLESLSRVSEADSYDV-----LI 154  
QY 148 MPGIDMMAGALHSRTAQLPQVRIDRPESLLGRSTTECIKSGVFWGVV-KQIGGLVDAIRG 206  
Db 155 LDEVMSVIGQLYSPTMR-----RLSAVDSLRLNRCSQIIAMDATVNSQFIDLISGLRG 210  
QY 207 D-----LVRDFGESFVEVITGNSRIIVPEIGPVSVVIDELAVL 245  
Db 211 DENIHTIVCTYAGVGFSGRTCTILRDMGIDTLVRVVKRS-----PEHEDVRTIHQLRGT 264  
QY 246 RGSDDLRLM 254  
Db 265 FFDELALRL 273

RESULT 9  
MREB\_HAEIN  
ID MREB\_HAEIN STANDARD; PRT; 351 AA.  
AC P44474;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Rod shape-determining protein mreB.  
GN MREB OR HI0037.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-Rd / KW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
Rd.";  
RL Science 269:496-512(1995).  
CC -!- FUNCTION: INVOLVED IN FORMATION OF THE ROD SHAPE OF THE CELL

CC (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE FTSA/MREB FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----

DR EMBL; U32689; AAC21715.1; ALT\_INIT.  
DR TIGR; HI0037; -  
DR InterPro; IPR001023; Hsp70.  
DR InterPro; IPR004753; MreB\_Mrl.  
DR PRODom; PD000089; Hsp70; 1.  
DR TIGRFAMs; TIGR00904; mreB; 1.  
KW Cell shape; Complete proteome.  
SQ SEQUENCE 351 AA; 37616 MW; D97CA0686403B279 CRC64;

Query Match 6.5%; Score 85; DB 1; Length 351;  
Best Local Similarity 23.0%; Pred. No. 3.3;  
Matches 50; Conservative 35; Mismatches 66; Indels 66; Gaps 10;

QY 52 EPQAIACSVVPSATAVGSALLE-----SLESVPVLTICCKLRFPRLDYATPHTFGADRL- 107  
Db 144 EPMAAAGAKLPVSTAVGSMVIDIGGGTTEVAVISL-----NGIVSSSVRIGGDRFD 196  
QY 108 -ALCAWSRHLFSE---KPV---IAVDIGTAITFDVLTGVNRYGGLIMPIDMMAGALHS 160  
Db 197 EAIISYVRRTFGSGVIGEPTAEIRIKQESAYIQE-----GDEIKEMEVGHNLAEAPRS 251  
QY 161 RTAQLPQVRIDRPESLLGRSTTECIKSGVFWGVVQKIGGLVDAIRGDLVRDFGESTVEV- 219  
Db 252 FT-----LTSRDVLEAIQ-----QPLNGIVAARVTALEECQPEHAADIF 290  
QY 220 ----IVTGGNSRIIVPEIGPVSVVIDELAVLRGSDLL 252  
Db 291 ERGMVLTGGG-----ALLRNIDILL 310

RESULT 10  
SYFB\_SYNY3  
ID SYFB\_SYNY3 STANDARD; PRT; 810 AA.  
AC P74296;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Phenylalanyl-tRNA synthetase beta chain (EC 6.1.1.20) (Phenylalanine--  
DE tRNA ligase beta chain) (PHERS).  
GN PHET OR SLL1553.  
OS Synechocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
OX NCBI\_TaxID=1148;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97061201; PubMed=8905231;  
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asanizu E., Nakamura Y.,  
RA Miyajima N., Hirose M., Sugita M., Sugita M., Sasamoto S., Kimura T.,  
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,  
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,  
RA Yamada M., Yasuda M., Tabata S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the  
RT entire genome and assignment of potential protein-coding regions.";  
RL DNA Res. 3:109-136(1996).  
CC -!- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) -> AMP +  
CC diphosphate + L-phenylalanyl-tRNA(Phe).  
CC -!- SUBUNIT: Tetramer of two alpha and two beta chains (BY  
CC SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- SIMILARITY: BELONGS TO THE PHENYLALANYL-TRNA SYNTHETASE BETA  
CC CHAIN FAMILY. SUBFAMILY 1.





RC STRAIN-ATCC 25085;  
RX MEDLINE-95331308; PubMed-7607244;  
RA Mueller U., Buckel W.;  
RT "Activation of (R)-2-hydroxyglutaryl-CoA dehydratase from  
RL Acidaminococcus fermentans";  
RN Eur. J. Biochem. 230:698-704(1995).  
RP [4]  
RX CHARACTERIZATION.  
RX MEDLINE-20558245; PubMed-11106419;  
RA Hans M., Buckel W., Bill E.;  
RT "The iron-sulfur clusters in 2-hydroxyglutaryl-CoA dehydratase from  
RT Acidaminococcus fermentans. Biochemical and spectroscopic  
RT investigations.";  
RL Eur. J. Biochem. 267:7082-7093(2000).  
RN [5]  
RX X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).  
RX MEDLINE-21140179; PubMed-11243821;  
RA Locher K.P., Hans M., Yeh A.P., Schmid B., Buckel W., Rees D.C.;  
RT "Crystal structure of the Acidaminococcus fermentans 2-  
RT hydroxyglutaryl-CoA dehydratase component A.";  
RN J. Mol. Biol. 307:297-308(2001).  
CC -!- FUNCTION: REQUIRED FOR THE ACTIVATION OF (R)-2-HYDROXYGLUTARYL-COA  
CC DEHYDRATASE. THIS PROTEIN IS EXTREMELY SENSITIVE TOWARDS OXYGEN.  
CC -!- COFACTOR: BINDS ONE 4FE-4S CLUSTER PER DIMER.  
CC -!- SUBUNIT: HOMODIMER.  
CC -!- SIMILARITY: TO E.COLI YJIL AND M.JANNASCHII MJ0004 AND MJ0800.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X59645; CAA42196.1; -.  
DR EMBL; X14252; CAA32464.1; -.  
DR PIR; S04476; S04476.  
DR PDB; 1HUX; 21-MAR-01.  
DR InterPro; IPR002731; ATPase\_BadF.  
DR Pfam; PF01869; Bcrad\_BadFG; 1.  
DR TIGRfam; TIGR00241; COA\_E\_activ; 1.  
DR ProDom; PD006344; ATPase\_BadF; 1.  
KW Iron-sulfur; 4Fe-4S; 3D-structure.  
FT METAL 127 127 IRON-SULFUR (4FE-4S).  
FT METAL 166 166 IRON-SULFUR (4FE-4S).  
FT CONFLICT 196 196 A -> P (IN REF. 2).  
FT CONFLICT 209 209 V -> L (IN REF. 2).  
FT CONFLICT 214 224 MTGVAQNYGV -> HDRRCSPWL (IN REF. 2).  
J SEQUENCE 260 AA; 27269 MW; 7E97044DBE805AC CRC64;  
  
Query Match 6.2%; Score 81; DB 1; Length 260;  
Best Local Similarity 20.7%; Pred. No. 5.2;  
Matches 51; Conservative 38; Mismatches 99; Indels 58; Gaps 10;  
  
QY 3 LVVDIGNSTTLAFTGDEEPSVESVPSALFADSDSTMREVFGNMARKHGEPQAIACSVV 62  
Db 6 LGIDVGSTASKCIILKDGKEIVAKSLVAVGTGTSPARSISEVLENAHMKEDMAF---- 61  
  
QY 63 PSATAVGSALLESFVSPVLTICCKLRFPPRLDYATPHTFGADRLALCAWSR-HLF----- 117  
Db 62 TLATGYGRNSLEGIAADKQMSLSC-----HAMGASFI---WPNVHTVIDIG 104  
  
QY 118 -SEKPIAVDIGTAITFDVLDVTGVNRYGGLIMPGIDMMAGALHSRTAQLPQVRIDRPE-- 174  
Db 105 QDVKVIHVENGTMNFQMDKCAAGTGRF-----LDVMANILEVKVSDLAELGAKSTKRV 160  
  
QY 175 -----SLGRSTTEC-IKSGVFWGVVQIGGLVDAIRGDLVRDFGESTVE 218  
Db 161 AISSTCTVFAESEVISQLSKGTDKIDIIAGIHRSVASRVIGL--ANRVGIKVD----- 211  
  
QY 219 VIVTGG 224  
|:|:|

Db 212 VVMTGG 217  
  
RESULT 13  
ACLY\_CAEEL STANDARD; PRT; 1106 AA.  
AC P53585;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Probable ATP-citrate (pro-S)-lyase (EC 4.1.3.8) (Citrate cleavage  
DE enzyme).  
GN D1005.1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-Bristol N2;  
RA Waterston R.;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: ATP CITRATE-LYASE IS THE PRIMARY ENZYME RESPONSIBLE FOR  
CC THE SYNTHESIS OF CYTOSOLIC ACETYL-COA IN MANY TISSUES (BY  
CC SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: ATP + citrate + CoA - ADP + phosphate +  
CC acetyl-CoA + oxaloacetate.  
CC -!- SUBUNIT: Homotetramer (By similarity).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -!- SIMILARITY: TO THE ALPHA AND BETA CHAINS OF SUCCINYL-COA SYNTHASES  
CC AND MALATE--COA LIGASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U58727; AAB00585.1; -.  
DR HSSP; P07459; 1SCU.  
DR WormPep; D1005.1; CE06997.  
DR InterPro; IPR003781; COA\_binding.  
DR InterPro; IPR000303; COA\_ligase.  
DR Pfam; PF00549; ligase-CoA; 1.  
DR Pfam; PF02629; CoA-binding; 1.  
DR PROSITE; PS01216; SUCCINYL\_COA\_LIG\_1; 1.  
DR PROSITE; PS00399; SUCCINYL\_COA\_LIG\_2; 1.  
DR PROSITE; PS01217; SUCCINYL\_COA\_LIG\_3; 1.  
KW Hypothetical protein; Lyase; Lipid synthesis; Phosphorylation;  
KW ATP-binding; Magnesium.  
FT ACT\_SITE 760 760 PHOSPHORYLATED IN THE COURSE OF  
FT ACT\_SITE 760 760 CATALYSIS (BY SIMILARITY).  
SQ SEQUENCE 1106 AA; 121616 MW; 49BEE17983BDEB3F CRC64;  
  
Query Match 6.2%; Score 81; DB 1; Length 1106;  
Best Local Similarity 26.6%; Pred. No. 28;  
Matches 38; Conservative 25; Mismatches 68; Indels 12; Gaps 5;  
  
QY 90 PFERLDYATPHTFGADRLALCAWS--RHLFESEKPIAVDIGTAITFDVLDVTGVNRYGGLI 147  
Db 531 YPFTGDNKQKYFGQKEILLPAYKSMKAFATHPDASIMVTFASMSVFETV---LEALE 587  
  
QY 148 MPGIDMMA---GALHSRTAQLPQVRIDRPESSLGRSTTECIKSGVFWGVVQIGGLVDA 203  
Db 588 FPQIKVIAIIAEGVPENQTRKLLKIAHDRGVTLVGPATVGGIKPGCF--KIGNTGGMMDN 645  
  
QY 204 IRGDLVRDFGESTVEIVTGGNS 226  
Db 646 ILASKLYRPG-SVAYVSRSGMS 667





Search completed: June 24, 2003, 22:11:51  
Job time : 6.16979 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: June 24, 2003, 22:08:46 ; Search time 11.9453 Seconds  
(without alignments)  
2373.336 Million cell updates/sec

Title: US-09-813-453A-11  
Perfect score: 1331  
Sequence: 1 MNKPLLSELIIDIGNTSIAF.....HLTVEGVRILGNSIDFKFVN 262

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues  
Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA.\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1331	100.0	262	9	US-09-813-453A-11
2	282	21.2	256	9	US-09-813-453A-55
3	276.5	20.8	262	9	US-09-813-453A-45
4	275.5	20.7	254	9	US-09-813-453A-47
5	271.5	20.4	258	9	US-09-813-453A-2
6	266.5	20.0	255	9	US-09-813-453A-7
7	262	19.7	250	9	US-09-813-453A-3
8	256.5	19.3	258	9	US-09-813-453A-49
9	231.5	17.4	233	9	US-09-813-453A-17
10	203	15.3	246	9	US-09-813-453A-9
11	197.5	14.8	258	9	US-09-813-453A-6
12	190	14.3	219	9	US-09-813-453A-57
13	189.5	14.2	260	9	US-09-813-453A-51
14	181.5	13.6	265	9	US-09-813-453A-4
15	170	12.8	273	9	US-09-813-453A-10
16	166.5	12.5	257	9	US-09-813-453A-53
17	159	11.9	262	9	US-09-813-453A-8
18	141.5	10.6	229	9	US-09-813-453A-12
19	138	10.4	272	9	US-09-712-363-276

20	138	10.4	272	9	US-09-813-453A-5	Sequence 5, Appl
21	136	10.2	241	9	US-09-813-453A-63	Sequence 63, Appl
22	135	10.1	244	9	US-09-813-453A-41	Sequence 41, Appl
23	121.5	9.1	209	9	US-09-813-453A-21	Sequence 21, Appl
24	121.5	9.1	476	10	US-09-774-414-3	Sequence 3, Appl
25	111	8.3	212	9	US-09-813-453A-59	Sequence 59, Appl
26	100	7.5	592	9	US-09-813-453A-43	Sequence 43, Appl
27	99.5	7.5	410	10	US-09-845-335-2	Sequence 2, Appl
28	96	7.2	592	9	US-09-813-453A-22	Sequence 22, Appl
29	93.5	7.0	242	9	US-09-813-453A-65	Sequence 65, Appl
30	92	6.9	460	9	US-09-813-453A-39	Sequence 39, Appl
31	92	6.9	1161	9	US-10-170-102-4	Sequence 4, Appl
32	90	6.8	257	9	US-09-813-453A-13	Sequence 13, Appl
33	88.5	6.6	223	9	US-09-895-913A-74	Sequence 74, Appl
34	88.5	6.6	223	9	US-09-813-453A-14	Sequence 14, Appl
35	88.5	6.6	223	9	US-09-813-453A-67	Sequence 67, Appl
36	88.5	6.6	370	9	US-10-106-698-6268	Sequence 6268, Ap
37	86.5	6.5	385	10	US-09-815-242-5713	Sequence 5713, Ap
38	86.5	6.5	518	10	US-09-815-242-12473	Sequence 12473, A
39	85.5	6.4	284	9	US-10-284-986-1	Sequence 1, Appl
40	85.5	6.4	284	10	US-09-846-808-1	Sequence 84, Appl
41	84.5	6.3	381	9	US-09-870-759-84	Sequence 2, Appl
42	84.5	6.3	1536	9	US-10-092-880-2	Sequence 5506, Ap
43	84	6.3	401	10	US-09-815-242-5506	Sequence 12261, A
44	84	6.3	401	10	US-09-815-242-12261	Sequence 20, Appl
45	83.5	6.3	248	9	US-09-813-453A-20	

ALIGNMENTS

RESULT 1

US-09-813-453A-11  
; Sequence 11, Application US/09813453A  
; Patent No. US20020168681A1

; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn ver. 2.0  
; SEQ ID NO 11  
; LENGTH: 262  
; TYPE: PRT  
; ORGANISM: Borrelia burgdorferi

US-09-813-453A-11

Query Match 100.0%; Score 1331; DB 9; Length 262;  
Best Local Similarity 100.0%; Pred. No. 5.1e-115;  
Matches 262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MNKP	LLSE	LIID	IGNTS	IAF	ALF	KDN	QVNL	FI	KM	KN	LM	LR	YDE	VY	SFFE	NF	DF	NKV	60	
Db	1	MNKP	LLSE	LIID	IGNTS	IAF	ALF	KDN	QVNL	FI	KM	KN	LM	LR	YDE	VY	SFFE	NF	DF	NKV	60	
QY	61	FIS	SV	PIL	NET	FK	NV	IF	SFF	KIK	PL	FI	GD	LN	YDL	T	FN	PK	YSD	KFL	120	
Db	61	FIS	SV	PIL	NET	FK	NV	IF	SFF	KIK	PL	FI	GD	LN	YDL	T	FN	PK	YSD	KFL	120	
QY	121	EN	Y	S	F	E	N	V	L	V	D	L	G	T	A	C	T	I	F	A	V	180
Db	121	EN	Y	S	F	E	N	V	L	V	D	L	G	T	A	C	T	I	F	A	V	180
QY	181	N	L	L	E	R	T	T	S	G	S	V	N	S	G	L	F	Y	Q	K	Y	240

Db 181 NLLERTTSGSVNSGLFYQYKYLIEGVYRDIKQMYKKKFNLIITGGNADLILSLIEIEFIF 240  
QY 241 NIHLTVEGVRIILGNSIDFKFVN 262  
Db 241 NIHLTVEGVRIILGNSIDFKFVN 262

RESULT 2  
US-09-813-453A-55  
; Sequence 55, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 55  
; LENGTH: 256  
; TYPE: PRT  
; ORGANISM: Clostridium difficile  
US-09-813-453A-55

Query Match 21.2%; Score 282; DB 9; Length 256;  
Best Local Similarity 31.3%; Pred. No. 2.6e-18;  
Matches 82; Conservative 54; Mismatches 94; Indels 32; Gaps 10;  
QY 9 LIIDIGNTSIAFALFKDNQVNLFIKMTNLMRLRYDE---VYSFFE-ENFDF-NVNKVFI 62  
Db 3 LVFDVGNNTMVLGIYKGDKLNVYWRIKTRDKTSDEYGILISNLFEDYDNVNISDIDV7II 62  
QY 63 SSVVPILNETFKNVIFSEFFKIKPLFI-----GFDLNYDLTFNPKSKDKFLLGSDVFANL 116  
Db 63 SSVVPNVMSHLENFCIKYCKKQPLIVGPIGIKTGLNLIKVD---NPKQ-----VGADRIVNA 114  
QY 117 VAAIENYSFENVLVVDLGTACTIFAVSRQDGIILGGINSGPLINFNSLLDNAYLIKFKPI 176  
Db 115 VAGIEKYGAPSIL-VDFGTATTFCATSEKGEYLGSTIAPGKISSEALFQSASKLPRVEL 173  
QY 177 STPNLLERTTSGSVNSGLFYQYKYLIEGVYRDIKQMYKKFN-----LIITGGNADLIL 231  
3 AKPGMTICKSTVSAMQSGIYGYVGLVD---KIIISIMKKELNCDVVKVIATGGLAKLIA 229  
QY 232 SLIE-IEFIFNIHLTVEGVRIIL 252  
Db 230 SETKSIDYVDGF-LTLEGLRII 250

RESULT 3  
US-09-813-453A-45  
; Sequence 45, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 45  
; LENGTH: 262  
; TYPE: PRT  
; ORGANISM: Bacillus anthracis  
US-09-813-453A-45  
Query Match 20.8%; Score 276.5; DB 9; Length 262;  
Best Local Similarity 27.8%; Pred. No. 8.6e-18;  
Matches 71; Conservative 63; Mismatches 100; Indels 21; Gaps 7;  
QY 10 LIIDIGNTSIAFALFKDNQVNLFIKMTNLMRLRYDE---VYSFFE-ENFDF-NVNKVFI 63  
Db 4 VLDVGNNTNAVLGVEEGELRQHWRMETDRHKTEDEYGMVLVKQLLEHGLSFEDVKGIIVS 63  
QY 64 SSVVPILNETFKNVIFSEFFKIKPLFI-----GFDLNYDLTFNPKSKDKFLLGSDVFANLV 117  
Db 64 SSVPPIMFALERMCEKFKIKPLVVGPGIKTGLNLIKVE---NPRE-----VGADRIVNA 115  
QY 118 AAIENYSFENVLVVDLGTACTIFAVSRQDGIILGGINSGPLINFNSLLDNAYLIKFKPI 177  
Db 116 AGIHLYG-SPLIIVDFGTATTTCYINEEKHYMGVITPIMISAEALYSRAAKLPRIEIT 174  
QY 178 TPNNLLERTTSGSVNSGLFYQYKYLIEGVYRDIKQMYKKFNLIITGGNADLILSLIEIE 237  
Db 175 KPSSVVGKNTVSAMQSGIYGYVGVQVEGIVKRMKEAKQEPKVIATGGLAKLISEESNVI 234  
QY 238 FIFNIHLTVEGVRIIL 252  
Db 235 DVVDPFLLKGLYML 249

RESULT 4  
US-09-813-453A-47  
; Sequence 47, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0.  
; SEQ ID NO 47  
; LENGTH: 254  
; TYPE: PRT  
; ORGANISM: Bacillus halodurans  
US-09-813-453A-47

Query Match 20.7%; Score 275.5; DB 9; Length 254;  
Best Local Similarity 28.1%; Pred. No. 1e-17;  
Matches 72; Conservative 58; Mismatches 105; Indels 21; Gaps 5;  
QY 9 LIIDIGNTSIAFALFKDNQVNLFIKMTNLMRLRYDEVYSFFEENFDF-----NVNKVFI 62  
Db 3 LVIDVGNTNTVLGVYQDETLLVHHWRLATSRQKTEDEYAMTVRSFLDHAGLQFQDIDGIVI 62  
QY 63 SSVVPILNETFKNVIFSEFFKIKPLFI-----GFDLNYDLTFNPKSKDKFLLGSDVFANL 116  
Db 63 SSVVPPMMFSLQCMCKKYFHTVPMIIGPGIKTGLNLIKVD---NPKQ-----VGADRIVNA 114  
QY 117 VAAIENYSFENVLVVDLGTACTIFAVSRQDGIILGGINSGPLINFNSLLDNAYLIKFKPI 176  
Db 115 VAAIELYGYP-AIVVDFGTATTTCYLINEKKQYAGGVIAPGIMISTEALYHRASKLPRIE 173  
QY 177 STPNLLERTTSGSVNSGLFYQYKYLIEGVYRDIKQMYKKFNLIITGGNADLILSLIEI 236







Query Match	14.2%;	Score 189.5;	DB 9;	Length 260;
Best Local Similarity	26.9%;	Pred. No. 9e-10;		
Matches 67;	Conservative 41;	Mismatches 130;	Indels 11;	Gaps 4

  

QY	9	LIIDIGNTSIAFALFKDNQVNLFIKMKTNLMRLRYDEVYSFFEEENFDEN-----VNKVF	52
		:     :    :	
Db	3	LAIEQGNNTMTFAIHDCASWAQWRSATESTRTADEYVVWLSQLLSMQGLGFRAIDAVII	62
		:     :    :	
QY	63	SSVVPILNETKVNIFSEFKIKPLFIGFDLNYDLTFNPYKSKDKFLLGSDVFANLVA	122
		: : : :     :   :   :   :   :   :   :   :   :   :   :   :	
Db	63	SSVVPQSIFNLRLNSRRYFNVEPLVIGENAKLGIDVRIEKPSE--ACADRLVNAIGAAMV	120
		: : : :     :   :   :   :   :   :   :   :   :   :   :   :	
QY	123	YSFENVLVDLGTACTIFAVSRQDGIILGGINSGPLINFNSLLDNAYLIKFKFPIS	180
		: : :           : :           : : :   :   :   :   :   :   :	
Db	121	YP-GPLVVIDSGTATFDIVAADGAFEGGIIAPGINLSMQALHEAAKLPRIAIORPAGN	179
		: : :           : :           : : :   :   :   :   :   :   :	
QY	181	NLLERTTSGSVNSGLEFYQKYKYLIEGVYRDIKQYKKKFNLIITCGNADLILSLIE	240
		: :   : :     :         :     : :           :       :	
Db	180	RIVGTDTVSAMQSGVFWGYISLIEGLVARIKAERGEPMVTIATGGVASLFEGATDS	239
		: :   : :     :         :     : :           :       :	
QY	241	NIHLTVEGV	249
		:   :   :	
Db	240	DSDLTIRGL	248
		:   :   :	

  

RESULT	14
--------	----





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:35:06 ; Search time 21.3688 Seconds  
(without alignments)  
2526.317 Million cell updates/sec

Title: US-09-813-453A-11  
Perfect score: 1331  
Sequence: 1 MNKPLLSLELIIDIGNTSIAF.....HLTVEGVRILGNSIDFKFVN 262

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

---Searched: 671580 seqs, 206047115 residues 671580  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_21:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archheap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1331	100.0	262	16 O51477	O51477 borrelia bu
2	293.5	22.1	255	16 Q8R7M2	Q8R7M2 thermoanaer
3	289.5	21.8	273	16 Q97EB4	Q97EB4 clostridium
4	275.5	20.7	254	16 Q9KGH5	Q9KGH5 bacillus ha
5	253.5	19.0	259	16 Q92F54	Q92F54 listeria in
6	248.5	18.7	259	16 Q8YAC5	Q8YAC5 listeria mo
7	248.5	18.7	259	16 Q8XHL5	Q8XHL5 clostridium
8	242.5	18.2	258	2 Q9F985	Q9F985 bacillus st
9	232.5	17.5	256	16 Q8RFE4	Q8RFE4 fusobacteri
10	203	15.3	246	16 Q9WZY5	Q9WZY5 thermotoga
11	189.5	14.2	261	16 Q9A6Z1	Q9A6Z1 caulobacter
12	181.5	13.6	265	16 Q9X8N6	Q9X8N6 streptomyc
13	170	12.8	273	16 Q83446	Q83446 treponema p
14	159	11.9	262	16 Q9RX54	Q9RX54 deinococcus
15	152.5	11.5	224	16 Q98Q93	Q98Q93 mycoplasma
16	143	10.7	274	16 Q9CD56	Q9CD56 mycobacteri

17	141.5	10.6	229	16 O67753	O67753 aquifex aeo
18	138	10.4	272	16 O06282	O06282 mycobacteri
19	121.5	9.1	209	16 Q9PIA9	Q9PIA9 campylobact
20	121.5	9.1	476	8 Q35809	Q35809 saccharomyc
21	119	8.9	517	16 Q8XMY0	Q8XMY0 clostridium
22	118	8.9	276	16 Q8YQD7	Q8YQD7 anabaena sp
23	118	8.9	625	10 Q98S73	Q98S73 guillardia
24	118	8.9	765	8 Q9MTD7	Q9MTD7 toxoplasma
25	116.5	8.8	553	17 Q975L9	Q975L9 sulfolobus
26	116	8.7	2178	2 Q46149	Q46149 clostridium
27	114	8.6	815	12 Q9EN43	Q9EN43 ansacta moo
28	112.5	8.5	963	16 Q98R23	Q98R23 mycoplasma
29	112	8.4	636	10 Q98RR1	Q98RR1 guillardia
30	111.5	8.4	629	12 Q9YW36	Q9YW36 melanoplus
31	111	8.3	212	2 Q32514	Q32514 desulfovibr
32	110.5	8.3	270	2 Q9AIY7	Q9AIY7 carsonella
33	110.5	8.3	1003	16 Q9PQ01	Q9PQ01 ureaplasma
34	110	8.3	654	12 Q9EN03	Q9EN03 ansacta moo
35	103.5	8.2	1179	5 Q8T2A2	Q8T2A2 dictyosteli
36	107.5	8.1	636	2 Q9AHW5	Q9AHW5 carsonella
37	107.5	8.1	1613	10 Q98S55	Q98S55 guillardia
38	107	8.0	477	5 Q97280	Q97280 plasmodium
39	107	8.0	590	10 Q98S67	Q98S67 guillardia
40	107	8.0	842	2 Q93U45	Q93U45 carsonella
41	104.5	7.9	267	2 Q9AI23	Q9AI23 carsonella
42	104.5	7.9	447	12 Q91ML6	Q91ML6 lumpy skin
43	104.5	7.9	448	12 Q9EMG8	Q9EMG8 ansacta moo
44	104	7.8	346	16 Q8RHE3	Q8RHE3 fusobacteri
45	104	7.8	660	16 Q8RGT2	Q8RGT2 fusobacteri

ALIGNMENTS

RESULT 1  
O51477 PRELIMINARY; PRT; 262 AA.  
ID O51477  
AC O51477;  
DT 01-JUN-1998 (TREMBlrel. 06, Created)  
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Hypothetical protein BB0527.  
GN BB0527.  
OS Borrelia burgdorferi (Lyme disease spirochete).  
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.  
OX NCBI\_TaxID=139;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 35210 / B31;  
RX MEDLINE=98065943; PubMed=9403685;  
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,  
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,  
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,  
RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,  
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,  
RA Utterback T., Wattney L., McDonald L., Artiach P., Bowman C.,  
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,  
RA Smith H.O., Venter J.C.;  
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia  
RT burgdorferi.";  
RL Nature 390:580-586(1997).  
DR EMBL; AE001154; AAC66882.1; --  
DR TIGR; BB0527; --  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMs; TIGR00671; baf; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 262 AA; 29822 MW; DC8CD558CC16B9AF CRC64;

Query Match 100.0%; Score 1331; DB 16; Length 262;  
Best Local Similarity 100.0%; Pred. NO. 2.9e-90;  
Matches 262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MNKPLLSLIIIDIGNTSIAFALFKDNQVNLFIKMKTNLMRLYDEVYSFFEENFDNFVNKV 60
Db 1 MNKPLLSLIIIDIGNTSIAFALFKDNQVNLFIKMKTNLMRLYDEVYSFFEENFDNFVNKV 60
QY 61 FISSVVPILNETFKNVIFSEFKIKPLFIQFDLNYDLTFNPKSKDKFLGSDVFANLVAAI 120
Db 61 FISSVVPILNETFKNVIFSEFKIKPLFIQFDLNYDLTFNPKSKDKFLGSDVFANLVAAI 120
QY 121 ENYSFENVLVVDLGTACTIFAVSRQDGILGGIINSGLPLNFNSLLDNAYLIKFKFPISTPN 180
Db 121 ENYSFENVLVVDLGTACTIFAVSRQDGILGGIINSGLPLNFNSLLDNAYLIKFKFPISTPN 180
QY 181 NLLERTTSGVNSGLFYQYKYLIEGVYRDIKQYKKKFNLIITGGNADLILSLIEIEFIF 240
Db 181 NLLERTTSGVNSGLFYQYKYLIEGVYRDIKQYKKKFNLIITGGNADLILSLIEIEFIF 240
QY 241 NIHLTVEGVRIILGNSIDFKFVN 262
Db 241 NIHLTVEGVRIILGNSIDFKFVN 262
RESULT 2
Q8R7M2
ID Q8R7M2 PRELIMINARY; PRT; 255 AA.
AC Q8R7M2;
DT 01-JUN-2002 (TremBLrel. 21, Created)
DT 01-JUN-2002 (TremBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Putative transcriptional regulator, homologs of Bvg accessory
DE factor.
GN TTE2381.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MB4T / JCM11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EMBL; AE013180; AAM25520.1; -
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 255 AA; 27816 MW; C3C620ECBC8CA6ED CRC64;
Query Match 22.1%; Score 293.5; DB 16; Length 255;
Best Local Similarity 29.7%; Pred. No. 4.4e-14;
Matches 76; Conservative 53; Mismatches 106; Indels 21; Gaps 5;
QY 9 LIIDIGNTSIAFALFKDNQVNLFIKMKTNLMRLYDEVYSFFEENFDNFN-----VNKVFI 62
Db 3 LAFDVGNTNIVMGVFKKLLHSFRISTDKNKTYDEYGMVLVNLIGYNGISLTDVVII 62
QY 63 SSVVPILNETFKNVIFSEFKIKPLFI-----GFDLNYDLTFNPKSKDKFLGSDVFANL 116
Db 63 SSVVPPLMNTLQVMSLKYFRTPKPIVVGPGIKTGINIKYD---NPKE-----VGADRIVNA 114
QY 117 VAAIENYSFENVLVVDLGTACTIFAVSRQDGILGGIINSGLPLNFNSLLDNAYLIKFKFPI 176
Db 115 VAAAYELYG-GPVIVIDFGTATTFCAISEKGEYLGGLIAPGLMISADALFQRTAKLPKIDL 173
QY 177 STPNLLERTTSGVNSGLFYQYKYLIEGVYRDIKQYKKKFNLIITGGNADLILSLIEI 236
Db 174 TKPPTVINRNTVASMQSGIYGHVGMVDYIVTRMKGEFAPSAYVVATGGFANMAEESKT 233
QY 237 EEFINHLTVEGVRIIL 252
Db 234 IDTVNEMLTLEGLRII 249
```

```
RESULT 3
Q97EB4
ID Q97EB4 PRELIMINARY; PRT; 273 AA.
AC Q97EB4;
DT 01-OCT-2001 (TremBLrel. 18, Created)
DT 01-OCT-2001 (TremBLrel. 18, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Predicted transcriptional regulator, homolog of Bvg accessory
DE factor.
GN CAC3200.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007815; AAK81136.1; -
DR InterPro; IPR004619; Baf.
DR InterPro; IPR000515; BPD_transp.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00671; baf; 1.
DR PROSITE; PS00402; BPD_TRANSF_INN_MEMBR; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 273 AA; 30331 MW; DE7B4D5923E72842 CRC64;
Query Match 21.8%; Score 289.5; DB 16; Length 273;
Best Local Similarity 29.6%; Pred. No. 9.5e-14;
Matches 77; Conservative 62; Mismatches 96; Indels 25; Gaps 8;
QY 9 LIIDIGNTSIAFALFKDNQVNLFIKMKTNLMRLYDE----VYSFTEEN-FDFN-VNKVFI 62
Db 14 LVLDVGNTNIVLGIYNDTKLTAEWRLSTDVLSRADEYGIQVMNLFQDDKLDPTLVEGVII 73
QY 63 SSVVPILNETFKNVIFSEFKIKPLFI-----GFDLNYDLTFNPKSKDKFLGSDVFANL 116
Db 74 SSVVPINIMYSLEHMIRKYFNPLVVGPGIKTGINIKYD---NPKE-----VGADRIVNA 125
QY 117 VAAIENYSFENVLVVDLGTACTIFAVSRQDGILGGIINSGLPLNFNSLLDNAYLIKFKFPI 176
Db 126 VAAHEIYK-RSLIIDFGTATTCFAVRENGDYLGAICPGIKVSSEALFEKAALPRVEL 184
QY 177 STPNLLERTTSGVNSGLFYQY----KYLIEGVYRDIKQYKKKFNLIITGGNADLILS 232
Db 185 IKPAYAICKNTISSIQSGIVYGVIGQVRYIVERMKEELOEGEREPLVVATGGLAKLISE 244
QY 233 LIEIEFIFNIHLTVEGVRIIL 252
Db 245 EAKNVDVINPFLTLEGLRII 264
RESULT 4
Q9KGH5
ID Q9KGH5 PRELIMINARY; PRT; 254 AA.
AC Q9KGH5;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Hypothetical protein BH0086.
GN BH0086.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
```







RA Fonstein M., Kyrpides N., Overbeek R.;  
RT "Genome sequence and analysis of the oral bacterium Fusobacterium  
RT nucleatum strain ATCC 25586.";  
RL J. Bacteriol. 184:2005-2018(2002).  
DR EMBL; AE010586; AAL94957.1; -.  
KW Complete proteome.  
SQ SEQUENCE 256 AA; 28601 MW; CDF681127F08914B CRC64;  
  
Query Match 17.5%; Score 232.5; DB 16; Length 256;  
Best Local Similarity 25.6%; Pred. No. 1.3e-09;  
Matches 64; Conservative 65; Mismatches 110; Indels 11; Gaps 5;  
  
QY 11 IDIGNTSIAFALEKDN-QVNLFIKMKTNMLRLRYDEVYSFFEEFDFN-----VNKVFI 63  
Db 5 IDIGNTHIVTGIYDNGELISTFRIATNDKMTDEYFSYFNNTKYNEISIKKVDAILIS 64  
  
QY 64 SVVPILNETKKNVIFSFFKIKPLFIGFDLNYDLTFNPKSKDFL-LGSDVVFANLVAAIEN 122  
Db 65 SVVPNIITFFQFFARKYKVEATIV--DLEKKLPFTFAKGINVTGFGADRIIDITEAMQK 122  
  
QY 123 YSFENVLVVDLGTACTIFAVSRQDGILGGIINSGPLINFNSLLDNAYLIKKFPISTPNNL 182  
Db 123 YPDKNLVIFDFGTA-TTYDVLKKGVIYGGILPGIDMSINALYGNATAKLPRVKFTTPSSV 181  
  
QY 183 LERTTSGSVNSGLFYQYKYLIEGVYRDIKQYKKKFNLIITGGNADLILSLIEIEFIFNI 242  
Db 182 LGDTMKQIQAAIFFGYAGQIKHIIKINEELNEEIFVLATGGLGKILSAEIDEIDYDA 241  
  
QY 243 HLTVEGVRIL 252  
Db 242 NLSLKGGLYTL 251  
  
RESULT 10  
Q9WZY5  
ID Q9WZY5 PRELIMINARY; PRT; 246 AA.  
AC Q9WZY5;  
DT 01-NOV-1999 (TremBLrel. 12, Created)  
DT 01-NOV-1999 (TremBLrel. 12, Last sequence update)  
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)  
DE Hypothetical protein TM0883.  
GN TM0883.  
OS Thermotoga maritima.  
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.  
OX NCBI\_TaxID=2336;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MSB / DSM 3109;  
RX MEDLINE=99287316; PubMed=10360571;  
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,  
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,  
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;  
RT "Evidence for lateral gene transfer between Archaea and Bacteria from  
RT genome sequence of Thermotoga maritima.";  
RL Nature 399:323-329(1999).  
DR EMBL; AE001754; AAD35964.1; -.  
DR TIGR; TM0883; -.  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMS; TIGR00671; baf; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 246 AA; 27154 MW; 9E0309AD462CF266 CRC64;  
  
Query Match 15.3%; Score 203; DB 16; Length 246;  
Best Local Similarity 24.6%; Pred. No. 1.9e-07;  
Matches 60; Conservative 62; Mismatches 112; Indels 10; Gaps 7;  
  
QY 9 LIIDIGNTSIAFALEKDNQVNLFIKMKTNMLRLRYDEVYSFFEE---NFDENVNVKVFISV 65  
Db 3 LLVDVGNTHSVFSITEDGKTRFRWRRLSTGVFQTEDELFSHLPLLGDAWREIKGIVASV 62

QY 66 VPILNETKKNVIFSFFKIKPLFIGFDLNYDLTFNPKSKDFLLGSDVVFANLVAAIENYSF 125  
Db 63 VPTQNTVIERFSQKYFPHISPIWVKAK-NGCVKWNVNKPSE--VGADRVANVAVFKEYG- 118  
  
QY 126 ENVLVVDLGTACTIFAVSRQDGILGGIINSGPLINFNSLLDNAYLIKKFPISTPNNL 185  
Db 119 KNGIIMGTATTVDLVV-NGSYEGGAILPGFFMMVHSLFRGTAKLPLVEVKPADVFVGK 177  
  
QY 186 TTSGSVNSGLFYQYKYLIEGVYRDIKQYKKKFNLIITGGNADLILSLIEIEFIFNIHLT 245  
Db 178 DTEENIRLGVVNGSVYALEGIIGRIKEY-GDLPVVLTTGGQSKIVKDMIKHE-IFDEDLT 235  
  
QY 246 VEGV 249  
Db 236 IKGV 239  
  
RESULT 11  
Q9A6Z1  
ID Q9A6Z1 PRELIMINARY; PRT; 261 AA.  
AC Q9A6Z1;  
DT 01-JUN-2001 (TremBLrel. 17, Created)  
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)  
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)  
DE Transcriptional activator, putative, Baf family.  
GN CC1935.  
OS Caulobacter crescentus.  
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
OC Caulobacter.  
OX NCBI\_TaxID=155892;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 19089 / CB15;  
RX MEDLINE=21173698; PubMed=11259647;  
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,  
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,  
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,  
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
RT "Complete genome sequence of Caulobacter crescentus.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
DR EMBL; AE005867; AAK23910.1; -.  
DR TIGR; CC1935; -.  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMS; TIGR00671; baf; 1.  
KW Complete proteome.  
SQ SEQUENCE 261 AA; 27965 MW; C19E60D7B0714EF5 CRC64;  
  
Query Match 14.2%; Score 189.5; DB 16; Length 261;  
Best Local Similarity 26.9%; Pred. No. 2e-06;  
Matches 67; Conservative 41; Mismatches 130; Indels 11; Gaps 4;  
  
QY 9 LIIDIGNTSIAFALEKDNQVNLFIKMKTNMLRLRYDEVYSFFEEFDFN-----VNKVFI 62  
Db 4 LAIEQGTNTMTFAIHGASVVAQWRSATESRTRTADEVYVWLSQLLSMOGLGFRDAVII 63  
  
QY 63 SSVVPILNETKKNVIFSFFKIKPLFIGFDLNYDLTFNPKSKDFLLGSDVVFANLVAAIEN 122  
Db 64 SSVVPQSIFNLRLNSRRYFNVEPLVIGENAKLIGIDVRIEKPSE--AGADRLVNAIGAAMV 121  
  
QY 123 YSFENVLVVDLGTACTIFAVSRQDGILGGIINSGPLINFNSLLDNAYLIKKFPISTP--N 180  
Db 122 YP-GPLVVIDSGTATFDIVAADGAFEGGIIAPGINLSMOALHEAAAKLPRIAIQRPAGN 180  
  
QY 181 NLLERTTSGSVNSGLFYQYKYLIEGVYRDIKQYKKKFNLIITGGNADLILSLIEIEFIF 240  
Db 181 RIVGTDTVSAMQSGVFWGYISLIEGLVARIKAEGERPEMTVIATGGVASLFEGATDSIDHF 240  
  
QY 241 NIHLTVEGV 249





**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:56 ; Search time 7.65383 Seconds  
(without alignments)  
1007.182 Million cell updates/sec

Title: US-09-813-453A-11  
Perfect score: 1331  
Sequence: 1 MNKPILLSLIDIGNTSIAF.....HLTVEGVRILGNSIDFKFVN 262

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	121.5	9.1	476	4	US-09-316-083-3	Sequence 3, Appli
2	99.5	7.5	410	4	US-09-140-466-2	Sequence 2, Appli
3	96.5	7.3	3200	2	US-08-477-451-8	Sequence 8, Appli
4	95	7.1	493	4	US-08-686-968C-4	Sequence 4, Appli
5	92.5	6.9	284	4	US-09-134-001C-2927	Sequence 2927, Ap
6	92	6.9	331	4	US-09-134-001C-3626	Sequence 3626, Ap
7	91.5	6.9	574	4	US-09-134-001C-5409	Sequence 5409, Ap
8	89	6.7	3079	5	PCT-US94-00198-4	Sequence 4, Appli
9	87	6.5	1854	4	US-09-004-838-108	Sequence 108, Appl
10	86	6.5	275	2	US-08-645-193B-19	Sequence 19, Appl
11	86	6.5	299	4	US-09-134-001C-3433	Sequence 3433, Ap
12	84.5	6.3	425	4	US-09-134-001C-5654	Sequence 5654, Ap
13	84.5	6.3	1095	4	US-09-206-942-69	Sequence 69, Appl
14	84.5	6.3	1536	1	US-08-038-682-2	Sequence 2, Appli
15	84.5	6.3	1536	1	US-08-302-832-2	Sequence 2, Appli
16	84.5	6.3	1536	2	US-08-530-198-2	Sequence 2, Appli
17	84.5	6.3	1536	2	US-08-469-880-2	Sequence 2, Appli
18	84.5	6.3	1536	2	US-08-728-470-2	Sequence 2, Appli
19	84.5	6.3	1536	2	US-08-617-697-2	Sequence 2, Appli
20	84.5	6.3	1536	4	US-08-719-641-2	Sequence 2, Appli
21	84.5	6.3	1536	4	US-09-206-942-67	Sequence 67, Appl
22	83	6.2	607	4	US-09-537-682-1	Sequence 1, Appli
23	83	6.2	873	3	US-09-187-331-6	Sequence 6, Appli
24	83	6.2	873	4	US-09-470-946-6	Sequence 6, Appli
25	83	6.2	873	4	US-09-438-906-2	Sequence 2, Appli
26	83	6.2	873	4	US-09-438-906-4	Sequence 4, Appli
27	83	6.2	925	2	US-08-392-946-1	Sequence 1, Appli

28	83	6.2	925	2	US-08-504-169-1	Sequence 1, Appli
29	83	6.2	925	5	PCT-US94-14893-1	Sequence 1, Appli
30	82.5	6.2	293	1	US-08-118-270-60	Sequence 60, Appl
31	82.5	6.2	293	5	PCT-US93-08528-60	Sequence 4, Appli
32	82.5	6.2	694	2	US-08-895-522-4	Sequence 4, Appli
33	82.5	6.2	694	3	US-09-195-391-4	Sequence 9, Appli
34	82	6.2	194	3	US-08-741-411-9	Sequence 9, Appli
35	82	6.2	218	2	US-08-531-525-19	Sequence 19, Appl
36	82	6.2	218	2	US-08-718-270A-19	Sequence 19, Appl
37	82	6.2	451	4	US-09-371-913A-9	Sequence 9, Appli
38	82	6.2	1180	4	US-09-206-942-65	Sequence 65, Appl
39	82	6.2	1188	4	US-09-206-942-63	Sequence 63, Appl
40	81	6.1	878	4	US-09-134-001C-4378	Sequence 4378, Ap
41	80.5	6.0	717	4	US-08-924-629C-5	Sequence 5, Appli
42	80.5	6.0	977	4	US-09-206-942-53	Sequence 53, Appl
43	80.5	6.0	983	4	US-09-206-942-51	Sequence 51, Appl
44	80.5	6.0	1565	6	5352450-2	Patent No. 5352450
45	79.5	6.0	220	1	US-07-991-867B-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1  
US-09-316-083-3  
; Sequence 3, Application US/09316083A  
; Patent No. 6280942  
; GENERAL INFORMATION:  
; APPLICANT: The Institute of Physical and Chemical Research  
; TITLE OF INVENTION: Endonuclease  
; FILE REFERENCE: PH-651  
; CURRENT APPLICATION NUMBER: US/09/316,083A  
; CURRENT FILING DATE: 1999-05-20  
; EARLIER APPLICATION NUMBER: JP98/141861  
; EARLIER FILING DATE: 1998-05-22  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 476  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-09-316-083-3

Query Match		9.1%;	Score 121.5;	DB 4;	Length 476;
Best Local Similarity		20.2%;	Pred. NO. 8.8e-05;		
Matches 73;		Conservative 55;	Mismatches 99;	Indels 135;	Gaps 19;
QY	15	NTSIAFALFKDNQVNLFIKMTNMLRYDE-----VYSFF----	49		
Db	46	NNMTCFIKWDNKKILLDDMYNVLYNHKQRTPMSKNRLMSKNIMDYKLLTYFYILN	105		
QY	50	-----EENFDENVKVFISVVPIILNETFKNVIFSFYFKIKPLFIGFDLNY-----DLT	97		
Db	106	KMKMEMDYNNNNNNISLK-----YNELLKNIM-----NNLNKTSNIETNLS	148		
QY	98	FNPKSKDKFLGSDV-FANLVAAI-ENYSFENV-----LVVDLGT---ACTIFAVS	143		
Db	149	NNFYLMCKYLINKYMKYLDMLNMPNNMENNINYGKLNKTKTVLDLNNNEFYDLSGLI	208		
QY	144	RQDGIL--GGI-----INSGPLN--FNSLLDNAYLIKFKPISTPNN	181		
Db	209	EGDGYIGPGGITITNHANDVLNTIFINKRIKNSILVEKWMDTLKDNPYFVNAFSINIKTN	268		
QY	182	LLERTTSGSVNSGLFYQYK-----YL-----IEGVYRDIK-----	211		
Db	269	LAKEKIFTNIYNKLYSDYKINQINNHIPYNYLNKLNKLPKNI-MDIKNNYWLAGTAA	327		
QY	212	-----QMYKKFNLIITGGNADLILSLIEI--EFIFNIH---LTVEGVRILGNS--ID	257		
Db	328	DGSFLSSMYPNPKDTLLFKNMRPSPVISQVETRKELIYLIQESFDLSISNVKKVGNRKLKD	387		
QY	258	FK 259			



Db 272 LKKNADPNIIITVGETCILTAINNHN-KNIIYKLLNYDIDINTIQNTLFKLEQD-----I 325  
QY 153 INSG-PLINFNSLLDN-----AYLIKFPPISTPNNLLERTTSGSVNSGLFYQYKYLI 203  
Db 326 INSTIDTYYYNNLVKKEFIKFLAYIVKRYEKNIGILFDYPTLG-----EYFV 375  
QY 204 EGVYRDIKQYKKKFNLIITGGNADLILSLIEIEFI 239  
Db 376 KFIDTCMMEIFEMKSD---KAGNTDIYSIIFTNKYI 408

RESULT 5  
US-09-134-001C-2927  
; Sequence 2927, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 2927  
; LENGTH: 284  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-2927

Query Match 6.9%; Score 92.5; DB 4; Length 284;  
Best Local Similarity 22.3%; Pred. No. 0.055;  
Matches 60; Conservative 35; Mismatches 83; Indels 91; Gaps 13;  
QY 8 ELIIDIGNTSIAFALFKDNQVNLFIKM-----KTNLML-----RYDEVYSFFFEEN 52  
Db 7 KLIGEVKMHYKFIKESKDN-TKLYMKVNDIQDAKANIIAHGVAEHLDRYDEITAY---- 61  
QY 53 FDFNVNKFVSSVVPILNETFKNVIFSFPKIKPLFIGFD-LNYDLTFNPKYKSDK--FLLG 109  
Db 62 -----LNEA-----GFSVIRYDQRGHRSEGRKRAFYSN 89  
QY 110 SDVFANLVAAIENY---SFENVLV---DLGTACTIFAVSRQDGIILGGIINSGPLINFNS 163  
Db 90 SNEIVEDLDAINVYKSNFEGKGYLVIGHSMGGYTVTLTGKHPNTVNGIITSGALTRYNN 149  
QY 164 LLDNAYLIKFPPISTPNNLLERTTSGSVNSGLFYQYKYKLIIEGVYRDIKQYKKKFNLIIT 223  
Db 150 KL-----FGNPDNRNISPD-----YIENNLSEGVCSLDLEVMEKYKLD----- 186  
QY 224 GGNADLILSLIEIEFIPIFNIHLTVEGVRIL 252  
Db 187 ----DLNAKOISMGLVFSI---MDGVRYL 208

RESULT 6  
US-09-134-001C-3626  
; Sequence 3626, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 3626  
; LENGTH: 331  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3626  
Query Match 6.9%; Score 92; DB 4; Length 331;  
Best Local Similarity 21.4%; Pred. No. 0.078;  
Matches 49; Conservative 31; Mismatches 85; Indels 64; Gaps 8;  
QY 32 IKMKTNMLRLRYDEVYSFFE-----ENFDENVNK--VFISSVVPILNETFKNV 76  
Db 6 VRMSNNIFTVYGEVPELVEKKTKEIVNDYLGQEIIDDYFNKYNLYESDLTPPIIETLTMP 65  
QY 77 IFS-----FFKIKPLFIGFDLNYDLTFNPKYKSDKFLGSDVFANLVAAIENYSFENVLV 131  
Db 66 FFSNKKAIIVKNSYVFTGEKFSKDLNHN---SDE-----LIKPLEKYDGENLIIF 112  
QY 132 DLGTACTIFAVSRQDGIILGGIINSGPLINFNSLLDNAYLIKFPPISTPNNLLERTTSGSV 191  
Db 113 EV-----YQPKLDER---KKLTKTLLKKNQALKKIEQMS 142  
QY 192 NSGLFYQYKYKLIIEGVYRDIKQYKKKFNLIITGGNADLILSLIEIEFIF 240  
Db 143 EKELKHWIKNTLNNNYKDIKQDALELF-IELTGVNYNIVSQELEKILIF 190

RESULT 7  
US-09-134-001C-5409  
; Sequence 5409, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 5409  
; LENGTH: 574  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-5409

Query Match 6.9%; Score 91.5; DB 4; Length 574;  
Best Local Similarity 30.5%; Pred. No. 0.2;  
Matches 36; Conservative 20; Mismatches 41; Indels 21; Gaps 7;  
QY 19 AFALFKDNQVNLFIKMKFNLMLRYD---EVYSFFEENFDENVNKFVFISSVVPILNETFK 74  
Db 435 ATVLEK---MHPFVRNKLNPAYISKYFLDISNYREVNVDVLFITDILISDYSSLIYE--- 488  
QY 75 NVIFSFFKIKPLFIGFDL-NYDLT---FNPKYKSDKFLGSDV--FANLVAAIENYSFE 126  
Db 489 ---FSVFKKPMFLFYAFDLEDIYTRDFEYEPYET--FVPGKIVKTFDELILALENDNDFE 541

RESULT 8  
PCT-US94-00198-4  
; Sequence 4, Application PC/TUS9400198  
; GENERAL INFORMATION:  
; APPLICANT: Schering Corp.  
; TITLE OF INVENTION: RAS Associated GAP Proteins  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Schering Corp.  
; STREET: 1 Giraldo Farms





GENERAL INFORMATION:  
APPLICANT: Kupke, Thomas  
APPLICANT: Gotz, Friedrich  
APPLICANT: Kempter, Christoph  
APPLICANT: Jung, Gunther  
TITLE OF INVENTION: Oxidative Decarboxylation of Peptides  
TITLE OF INVENTION: Catalyzed by Flavoprotein EpiD  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
STREET: 1100 New York Avenue, Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/645,193B  
FILING DATE: 13-MAY-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Esmond, Robert W.  
REGISTRATION NUMBER: 32,893  
REFERENCE/DOCKET NUMBER: 0652.1540000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 275 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
US-08-645-193B-19

Query Match 6.5%; Score 86; DB 2; Length 275;  
Best Local Similarity 20.5%; Pred. No. 0.26;  
Matches 55; Conservative 49; Mismatches 94; Indels 70; Gaps 13;  
QY 4 PLLSELIIDIGN-----TSIAFALFKDNQVNLFIKMTNMLRLRYDEV--YSFFE 50  
Db 28 PILIIVILPIGNIMKRVSSKSQEATAKLSSYSNRLSTIKLIKLTSTYNIKIKNYTLK 87  
QY 51 ENDFNVNKV-FISSVVPILNETFKNVIFSFFKIKPLFIGFDLNYDLTFNPKYKSDKFLLG 109  
Db 88 NIFDIELHKIKVLSFFPEIPIMNLILFINIFGI-----LFLGYLH-----ENNMMKS 133  
QY 110 SDVFA-----NLVAAIENYSFENVLVVDLGTACTIFAVSRQDGLGILGINSGLPIN 160  
Db 134 GDMFAYVLYLEQIINPIVSITSYWE-----VQRAIG-----SSDRILK 172  
QY 161 FNSLLDNAYLIKKEPISTPNLLERTTSGSVNSGLFYQYKYLIEGVYRDIKQMYKKKFNL 220  
Db 173 INKEPEVLTIK-----TTYNNFVQKMEINDLN--FTKDNKQIINSISLDLHKGY--IYNI 224  
QY 221 IITG--GNADL--ILSLIEIEFIFNIHL 244  
Db 225 IGESGCGKSTLLNLAGLNTYTGNIYL 252

RESULT 11  
US-09-134-001C-3433  
; Sequence 3433, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 3433  
LENGTH: 299  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3433

Query Match 6.5%; Score 86; DB 4; Length 299;  
Best Local Similarity 23.8%; Pred. No. 0.3;  
Matches 67; Conservative 50; Mismatches 87; Indels 78; Gaps 20;  
QY 32 IKMKTNMLRLRYDEVYSFFEENFDNVNK-----VFISSVVPILNETFKNVIFS 80  
Db 15 VKLKNLISELKVMRTIMH-----NINKWLLVPYLLWMVIFI--IIPVI-----LLVYFSF 64  
QY 81 FKIKPLFIGFDLNYDLTFNPKYKSDKFLGSDVFANLVAIE-----NYS-FENV 128  
Db 65 IDIHGF-SF-TNYEQVFST-RYLMKFIDSIIWYAALITMITLIISYPAAAYFISYRFRQNI 121  
QY 129 LVVDLGTACTIFAVSRQ---DGILG--GIINSG-----PLIN--FNS---LLDNAYLIK 172  
Db 122 LMLLIPTWINLLKTYAFIGLLGHDGVINQALHIFQIPKLNLLFTSGAFLLVASIYI 181  
QY 173 KPPI-----STPNLLERTTSGSVNSGLFYQYKYL-----EGVYRDIKQMYKKKN 219  
Db 132 PFMILPIFNSMKAIPNNILQ--ASNDLGASTFTTRKRVIVPLTREGIKTVGVTFIPALS 239  
QY 220 L-----IITGGNADLILSLIEIEFIFNIHLTVEGVRILGNSI 256  
Db 240 LFMITRLIAGNKVINVGTAIEEQF-----LTIQNYG-LGSTI 275

RESULT 12  
US-09-134-001C-5654  
; Sequence 5654, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 5654  
; LENGTH: 425  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-5654

Query Match 6.3%; Score 84.5; DB 4; Length 425;  
Best Local Similarity 21.9%; Pred. No. 0.72;  
Matches 37; Conservative 39; Mismatches 58; Indels 35; Gaps 9;

QY 18 IAFALFKDNQV----NLFIKMTNMLRLRYDEVYSFFEENFDNVNKFVFISSVVPILNETF 73  
Db 70 VAFSIDKIEQIYELNQLFTYQYENLKIMNKCSDVNVNQVIFNLNEKSI-----REIY 122  
QY 74 KNVIFSFFKIKPLFIGFDLNYDLT--FNPKYKSDKFLGSDVFANLVAAIENYSFENVLVV 131  
Db 123 RNIL---KIQNIELEYKIGLIDSCMFN-----DTAQPKSLASQIKRLKFD-YLYI 168

QY 132 DLGTACTIFAVSRQDGLGILGINSGLINFNLSLDNAYLIKPFISTPN 180  
Db 169 DNARLKSSYLLDNEGLL--LKN---ILQFKHLIDD--LKQDFDSSN 209

RESULT 13

US-09-206-942-69  
; Sequence 69, Application US/09206942  
; Patent No. 6432669

; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.

; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H.

; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High

; FILE REFERENCE: 1038-861 MIS:jb  
; CURRENT APPLICATION NUMBER: US/09/206,942

; EARLIER FILING DATE: 1998-12-08  
; EARLIER FILING DATE: 1998-10-07

; NUMBER OF SEQ ID NOS: 95  
; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 69  
; LENGTH: 1095

; TYPE: PRT  
; ORGANISM: Haemophilus influenzae

US-09-206-942-69

Query Match 6.3%; Score 84.5; DB 4; Length 1095;

Best Local Similarity 22.7%; Pred. No. 2.9;

Matches 64; Conservative 42; Mismatches 77; Indels 99; Gaps 16;

QY 23 FKDNQVNL-----FIKMTNLMRLRYDEVYSFFEENEDFNVNKVFISVVPIILNET-- 72  
Db 160 FRFNNVSLNGTSGLOFTTKRTN--KY-AITNKFEGTLNIS-GKVNISMVLP-KNESGY 213

QY 73 --FKNVIFSPFKIKPLFI-----GFDLNYDLTFNPK 102  
Db 214 DKFKG--RTYWNLTSLNVSEGEFNLTDISRGSDSAGTLTQPYNLNGISFNKDTTFNVER 271

QY 103 SDK--FLIGSDVFANLVAAIENYSFENVLVVDLTAC--TIFAVSRQDGLGILGINSGL 158  
Db 272 NARVNFEDIKAPIGINKYSSLNYSFNGNISVGGSGVDFTLLASSNVQTPGVVINS--- 328

QY 159 INFNSLLDNAYLIKPFISTPNLLERTTSGSVNSGLFYQYKYLIEGVYRDIKQMYKKKF 218  
Db 329 -----KYFNVSTGSSLRFK-TSGSTKTG-----FSIE-----KDL 357

Y 219 NLIITGGNADLILSLIEIE-----FIFNIHLTVEGVRI 251  
Jb 358 TLNATGGN-----ITLLQVEGTDGMIGKIVAKKNITFECCNI 395

RESULT 14

US-08-038-682-2

; Sequence 2, Application US/08038682  
; Patent No. 5549897

; GENERAL INFORMATION:  
; APPLICANT: BARENKAMP, STEPHEN J

; APPLICANT: ST. GEME III, JOSEPH W  
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS

; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS

; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Shoemaker and Mattare, Ltd  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

; CITY: Arlington  
; STATE: Virginia

; COUNTRY: U.S.A.  
; ZIP: 22202-0286

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/038,682

; FILING DATE: 16-MAR-1993  
; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:  
; NAME: BERKSTRESSER, JERRY W

; REGISTRATION NUMBER: 22,651  
; REFERENCE/DOCKET NUMBER: 1038-293

; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 415-0810

; TELEFAX: (703) 415-0813  
; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1536 amino acids

; TYPE: amino acid  
; STRANDEDNESS: single

; TOPOLOGY: linear  
; MOLECULE TYPE: protein

US-08-038-682-2

Query Match 6.3%; Score 84.5; DB 1; Length 1536;

Best Local Similarity 22.7%; Pred. No. 4.7;

Matches 64; Conservative 42; Mismatches 77; Indels 99; Gaps 16;

QY 23 FKDNQVNL-----FIKMTNLMRLRYDEVYSFFEENEDFNVNKVFISVVPIILNET-- 72  
Db 601 FRFNNVSLNGTSGLOFTTKRTN--KY-AITNKFEGTLNIS-GKVNISMVLP-KNESGY 654

QY 73 --FKNVIFSPFKIKPLFI-----GFDLNYDLTFNPK 102  
Db 655 DKFKG--RTYWNLTSLNVSEGEFNLTDISRGSDSAGTLTQPYNLNGISFNKDTTFNVER 712

QY 103 SDK--FLIGSDVFANLVAAIENYSFENVLVVDLTAC--TIFAVSRQDGLGILGINSGL 158  
Db 713 NARVNFEDIKAPIGINKYSSLNYSFNGNISVGGSGVDFTLLASSNVQTPGVVINS--- 769

QY 159 INFNSLLDNAYLIKPFISTPNLLERTTSGSVNSGLFYQYKYLIEGVYRDIKQMYKKKF 218  
Db 770 -----KYFNVSTGSSLRFK-TSGSTKTG-----FSIE-----KDL 798

QY 219 NLIITGGNADLILSLIEIE-----FIFNIHLTVEGVRI 251  
Db 799 TLNATGGN-----ITLLQVEGTDGMIGKIVAKKNITFECCNI 836

RESULT 15

US-08-302-832-2

; Sequence 2, Application US/08302832  
; Patent No. 5603938

; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J

; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus

; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Shoemaker and Mattare, Ltd.  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

; CITY: Arlington  
; STATE: Virginia

; COUNTRY: U.S.A.  
; ZIP: 22202-0286

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/302,832

; FILING DATE: 16-SEP-1994



**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:33:11 ; Search time 4.55691 Seconds  
(without alignments)  
2384.688 Million cell updates/sec

Title: US-09-813-453A-11  
Perfect score: 1331  
Sequence: 1 MNKPLLSLIIIDIGNTSIAF.....HLTVEGVRLGNSIDFKFVN 262

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues  
--total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	231.5	17.4	233	1 YACB_BACSU	P37564 bacillus su
2	120.5	9.1	476	1 RF3_SACBA	P05512 saccharomyc
3	109.5	8.2	2136	1 YCF2_MARPO	P09975 marchantia
4	108	8.1	311	1 PRMA_CLOAB	P45558 clostridium
5	106.5	8.0	519	1 YC82_ASTLO	P58145 astasia lon
6	102	7.7	1024	1 RPOB_PLAFA	P21421 plasmodium
7	100	7.5	355	1 Y198_RICPR	Q92dw7 rickettsia
8	99.5	7.5	516	1 YC82_EUGGR	P30397 euglena gra
9	99	7.4	971	1 YQ1A_CAEEL	Q09281 caenorhabdi
10	96	7.2	554	1 Y478_RICPR	Q92d66 rickettsia
11	95.5	7.2	773	1 SYFB_CAMJE	Q9pp35 campylobact
12	95	7.1	971	1 SEC5_YEAST	P89102 saccharomyc
13	93.5	7.0	507	1 MVIN_RICPR	Q9zwc4 rickettsia
14	93.5	7.0	1783	1 Y468_MYCGE	Q49460 mycoplasma
15	93	7.0	462	1 ANFK_AZOVI	P16267 azotobacter
16	93	7.0	1149	1 A8A1_BOVIN	Q29449 bos taurus
17	92.5	6.9	269	1 FLIR_BORBU	Q44907 borrelia bu
18	92.5	6.9	2054	1 YCF2_PINTH	P41653 pinus thunb
19	92	6.9	518	1 PAC2_YEAST	P39937 saccharomyc
20	92	6.9	1149	1 A8A1_MOUSE	P70704 mus musculu
21	92	6.9	1164	1 A8A1_HUMAN	Q9Y2q0 homo sapien
22	91.5	6.9	648	1 GCP4_DROME	Q9vku7 drosophila
23	91.5	6.9	1442	1 DPO3_UREPA	Q9pqb4 ureaplasma
24	91	6.8	322	1 GPT_SULSO	P96000 sulfolobus
25	91	6.8	363	1 V363_ASFB7	P23164 african swi
26	90.5	6.8	504	1 Y795_METJA	Q58205 methanococc
27	90.5	6.8	575	1 RPOC_PLAFA	P21422 plasmodium
28	90.5	6.8	602	1 EX5A_BUCAI	P57530 buchnera ap
29	90.5	6.8	1296	1 BXG_CLOBO	Q60393 clostridium
30	90	6.8	708	1 PTA_BUCAI	P57273 buchnera ap
31	90	6.8	795	1 SYFB_BUCAI	P57230 buchnera ap
32	89.5	6.7	279	1 PPNK_BORBU	O51291 borrelia bu
33	89.5	6.7	380	1 IPYR_PLAFA7	O77392 plasmodium

34	89.5	6.7	646	1 NTP1_HAEPV	O37319 heliothis a
35	89	6.7	339	1 RMAR_CANGA	P21358 candida gla
36	89	6.7	361	1 V227_FOWPV	Q9J508 fowlpox vir
37	89	6.7	623	1 PPID_BUCAI	P57550 buchnera ap
38	89	6.7	898	1 YMV6_YEAST	Q04748 saccharomyc
39	89	6.7	3079	1 IRA2_YEAST	P19158 saccharomyc
40	88.5	6.6	229	1 Y066_BORBU	O51093 borrelia bu
41	88.5	6.6	259	1 IF2C_GALSU	Q08810 galdieria s
42	88.5	6.6	461	1 TRPC_BUCSC	Q44603 buchnera ap
43	88	6.6	813	1 VGLH_HSVMD	P36336 marek's dis
44	88	6.6	1250	1 BXE_CLOBO	Q00496 clostridium
45	87.5	6.6	354	1 PON3_HUMAN	Q15166 homo sapien

ALIGNMENTS

RESULT 1  
YACB\_BACSU  
ID YACB\_BACSU STANDARD; PRT; 233 AA.  
AC P37564;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein yacB.  
GN YACB.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=96051385; PubMed=7584024;  
RA Ogasawara N., Nakai S., Yoshikawa H.;  
RT "Systematic sequencing of the 180 kilobase region of the Bacillus  
RT subtilis chromosome containing the replication origin.";  
RL DNA Res. 1:1-14(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Mael C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,  
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
RA Viari A., Wambutt R., Wedler E., Wedler H., Weltzenegger T.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
RT subtilis.";  
RL Nature 390:249-256(1997).  
CC -1- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.



```
DR InterPro; IPR003959; AAA_ATPase_cent.
DR Pfam; PF00004; AAA; 1.
KW Chloroplast; Hypothetical protein.
SQ SEQUENCE 2136 AA; 259911 MW; 5BD170C9CCF61197 CRC64;

Query Match      8.2%; Score 109.5; DB 1; Length 2136;
Best Local Similarity 23.4%; Pred. No. 1.3;
Matches 80; Conservative 55; Mismatches 98; Indels 109; Gaps 21;

QY 6 LSELIIDIGNT-----SIAFALF-----KDNQVNLFIKMTNLMRLRYDEV 45
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 235 INQLKIDIKNSKYNKNSIGFEVFLAFCEKLLFEVEFLSKPNNNL--QMKNCL- 286
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 46 YSFFEENFDEN-----VNKFVSISSVVPILNETFKNVIFSFKKIKPLFIGFDLN 93
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 287 -----ENFSFLDIFCILNKKLPVWVKKIFKN-LQNFNESDKKLIESFFLLKIKGNLYFKN 340
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 94 Y--DLTFNPYKSD-----KF--LLGSDVFANLVAAIEN--YSFENVLVVDLGFACITFAV 142
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 341 YIEFVTWQSKKDCDLDFNKFENLNNSIYIKIEELFSDIYIKFSKYILYEGKSKTIKQ 400
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 143 SRQDGILGGIINSGPLINFSNL-----LDNAYLIKKEPI--STPNNL 182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 401 SFNNNIYYKKLNS--IFNENTIFYFDSNNLLFDWLKKNYYINNKPKLSFLIYSSISNQF 458
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 183 L---ERTTSGSVNSGLFYQKYKYLIEGVYRD--IKQMYKKKFNLIITGGNADL-----ILS 232
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 459 ILFFKQKNSKSFN-----KNLVKKNSKDVITNVFSKENKIEINNFSSKIYIAFFEILS 511
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 233 LIEIE--FIFNIHLTVEGV-----RILGNSI-----DFKFN 262
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 512 INEIDNKFEIN-KISLKNINKKKQKRFYLNKIKSSDNFRFIN 552
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 4
PRMA_CLOAB      STANDARD;      PRT;      311 AA.
ID  P45558;
DT  01-NOV-1995 (Rel. 32, Created)
DT  16-OCT-2001 (Rel. 40, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Probable methyltransferase (EC 2.1.1.-).
GN  PRMA OR CAC1284.
OS  Clostridium acetobutylicum.
OC  Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC  Clostridium.
OX  NCBI_TaxID=1488;
N  [1]
P  SEQUENCE FROM N.A.
RC  STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX  MEDLINE=21359325; PubMed=11466286;
RA  Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA  Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA  Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA  Bennett G.N., Koonin E.V., Smith D.R.;
RT  "Genome sequence and comparative analysis of the solvent-producing
RT  bacterium Clostridium acetobutylicum.";
RL  J. Bacteriol. 183:4823-4838(2001).
RN  [2]
RP  SEQUENCE OF 1-298 FROM N.A.
RX  MEDLINE=94123950; PubMed=7507453;
RA  Behrens S., Narberhaus F., Bahl H.;
RT  "Cloning, nucleotide sequence and structural analysis of the
RT  Clostridium acetobutylicum dnaJ gene.";
RL  FEMS Microbiol. Lett. 114:53-60(1993).
RN  [3]
RP  REVISIONS.
RA  Behrens S.;
RL  Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC  -!- INDUCTION: BY HEAT SHOCK.
CC  -!- SIMILARITY: STRONG, TO ENTEROBACTERIAL RIBOSOMAL PROTEIN L11
CC  METHYLTRANSFERASE (PRMA).
CC  -!- SIMILARITY: TO OTHER METHYLTRANSFERASES.
```

```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE007641; AAK79255.1; -
DR EMBL; X69050; CAA48793.1; -
DR InterPro; IPR004498; PRMA.
DR InterPro; IPR000051; SAM_bind.
DR TIGRFAMS; TIGR00406; prmA; 1.
KW Heat shock; Transferase; Methyltransferase; Complete proteome.
SQ SEQUENCE 311 AA; 35028 MW; 2D71786B3C974922 CRC64;

Query Match      8.1%; Score 108; DB 1; Length 311;
Best Local Similarity 20.9%; Pred. No. 0.17;
Matches 58; Conservative 52; Mismatches 87; Indels 80; Gaps 15;

QY 8 ELIIDIGNTSIAFALFKDNQVNLFIKMTNLMRLRYDEVYSFFEENFD-----FNVN---K 59
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : |
Db 55 ESLNVKDGAVIKAYYKDDH-----NFDESVKYIERSIDKLSEFGINKGEGK 101
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : |
QY 60 VFISVVVPILNET-FKNVIFSEFK-----IKPLFIGF-DLNYDLTFNPYKSDKFL 108
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : |
Db 102 VFNKV-----NETDWNENKKYKPKIGARIVVVKPLWEEYTPKDYELMLNMDPGMAFGT 157
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : |
QY 109 GSDVFANL-VAAIENYSFENVLVVDLGTACTIFAVS-----RQDGILGGIINSGPL- 158
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : |
Db 158 GTHETTRMCIOALERYVNEDEAEVFDIGTSGILAIATAAKLNAAKVLGVLDSDVAVKAAKE 217
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : |
QY 159 -INFNSLLDNAYLIKPKFPSTPNLLERTTSG-----SVNSGLFYQ 198
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : |
Db 218 NIOYNN-VNNIEILH-----GNLME-VVQKGADIIVANIIADVINILIPDINKFLKTD 268
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : |
QY 199 KYLLIEGVYRD-----IKQMYKKKFNLIITGGNADLI 230
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 269 GYFISSGIKDRADVDVIENLKKKFEIEVNNQGEWI 305

RESULT 5
YC82_ASTLO
ID  YC82_ASTLO      STANDARD;      PRT;      519 AA.
AC  P58145;
DT  16-OCT-2001 (Rel. 40, Created)
DT  16-OCT-2001 (Rel. 40, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Hypothetical 65.0 kDa protein ycf82 (ORF519).
GN  YCF82.
OS  Astasia longa (Euglenophycean alga).
OG  Chloroplast.
OC  Eukaryota; Euglenozoa; Euglenida; Euglenales; Astasia.
OX  NCBI_TaxID=3037;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=CCAP 1204-17a;
RX  MEDLINE=21080522; PubMed=11212895;
RA  Gockel G., Hachtel W.;
RT  "Complete gene map of the plastid genome of the nonphotosynthetic
RT  euglenoid flagellate Astasia longa.";
RL  Protist 151:347-351(2000).
CC  -!- SIMILARITY: BELONGS TO THE YCF82 FAMILY. SOME SIMILARITY TO
CC  GROUP II INTRON MATURASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```



```
CC -----
DR EMBL; AJ294725; CAC24599.1; -.
KW Hypothetical protein; Chloroplast.
SQ SEQUENCE 519 AA; 64996 MW; CCD0903E3450F124 CRC64;

Query Match      8.0%; Score 106.5; DB 1; Length 519;
Best Local Similarity 25.3%; Pred. No. 0.42;
Matches 75; Conservative 40; Mismatches 96; Indels 85; Gaps 18;

QY 11 IDIGNTSIAFALEKDNQVNLFIKMKTNLMRLRYDEVYSFEE--ENFDENVNKVFISVVYP 67
Db 106 IEIINLSFKENFF-----LFYKFCFNNLVKF-VMYPFLELKTENFVFNFN--FYTNLYD 156
QY 68 ILNETFKNVIFSFFKIKPLFIGFD--LNY-----DLTFNP-----YKSDKFLLGSD 111
Db 157 IFLNFKILLKNFMKLYFTINFNNKFNMMKKWLYKNLTNFNINFLKYFLKLNNFIFNGS 216
QY 112 V-----FANLVAIENYSFENVLVDLGTACTIFAVSRQDGILGGIIN-----S 155
Db 217 ILFFIFNFMNGKSFQKCFMNNEFIYVSSLKNIFYFFFKSYNEFNICINLYIYFLKYK 276
QY 156 GPLINFNSLLDNAYLKKPPISTPNLLERTTSG-SVNSGLFYQ--YKYLIEGVYRDIKQ 212
Db 277 G--INFNFINDH---VKNE-----LTNGLTFNFIKFFNFYKFKFLKLNKDINY 320
QY 213 MYKKKFNLIITGGNADLILSLIEIEFIFNIHLTVEGVRIILGNSID-----FKFVN 262
Db 321 Y---KFN-----IKTLIKNFYIKNIFYT---ICLINNKIKNWLDKYFLFVN 360

RESULT 6
RPOB_PLAFA
ID RPOB_PLAFA STANDARD; PRT; 1024 AA.
AC P21421;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE DNA-directed RNA polymerase beta chain (EC 2.7.7.6).
GN RPOB.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BW(C10);
RX MEDLINE=95107345; PubMed=7808472;
RA Gardner M.J., Goldman N., Barnett P., Moore P.W., Rangachari K.,
RA Strath M., Whyte A., Williamson D.H., Wilson R.J.M.;
RA "Phylogenetic analysis of the rpoB gene from the plastid-like DNA of
Plasmodium falciparum.";
AL Mol. Biochem. Parasitol. 66:221-231(1994).
RN [2]
RP SEQUENCE OF 328-1024 FROM N.A.
RX MEDLINE=91187055; PubMed=2011147;
RA Gardner M.J., Williamson D.H., Wilson R.J.M.;
RA "A circular DNA in malaria parasites encodes an RNA polymerase like
that of prokaryotes and chloroplasts.";
RL Mol. Biochem. Parasitol. 44:115-124(1991).
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
SUBSTRATES.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
[RNA](N).
CC -!- MISCELLANEOUS: THIS RNA POLYMERASE IS ENCODED ON A CIRCULAR DNA.
CC -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.
CC -----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

```
CC -----
DR EMBL; X75544; CAA53232.1; -.
DR EMBL; X52177; CAA36427.1; -.
DR EMBL; X95275; CAA64572.1; -.
DR PIR; S10438; RNZQBF.
DR InterPro; IPR001572; RNA_pol_B.
DR Pfam; PF00562; RNA_pol_B; 1.
DR PROSITE; PS01166; RNA_POL_BETA; 1.
KW Transferase; Transcription; DNA-directed RNA polymerase.
SQ SEQUENCE 1024 AA; 122185 MW; B49FBDED7C9BD412 CRC64;

Query Match      7.7%; Score 102; DB 1; Length 1024;
Best Local Similarity 25.0%; Pred. No. 2.1;
Matches 67; Conservative 44; Mismatches 97; Indels 60; Gaps 16;

QY 18 IAFALFKDNQVNLFIKMKTNLM-LRYDEVYSFFFEENFDENVNK-----VF 61
Db 478 IPFIHYNDSIRNLSIKMHTQIVPIIYNLSNII-TNYNFILNKYLNHLIISYQEGIVY 536
QY 62 ISSVVPFILNETF-KNVIFSFFKIKPLFIGFDLNYDLTFNP--YKSDKFLLGSDVEANLVA 118
Db 537 VSCIKIIRDLFNRIIYYLNYYKKI----NQNILLYKPIVWVGKVNIGQILAINSNL 592
QY 119 AIENYSFENVLVVDLGTACTIFAVSRQDGILGGIINSGPLIN--FNSLLDNAYLKKPPI 176
Db 593 LNSEYSLGNLLVGYG--SYLGYEYEDAI---IISRKILYNNLYTSLHLNLYEIS---L 643
QY 177 STPNLLERTTSGSVNSGLFYQYKYLIEGVYRDIKQMYKKKFNLIITG-----GNADLILS 232
Db 644 NIINNIPEIC---SINLSKMY-----YKNIKHL--DKYGIKEGTYILANNILISK 689
QY 233 LIEIEFIFNIHLTVEGVRIILGNSIDFKF 260
Db 690 LMFMPFIFN-----NKSLINIINFLF 710

RESULT 7
Y198_RICPR
ID Y198_RICPR STANDARD; PRT; 355 AA.
AC Q9ZDW7;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein RP198 precursor.
GN RP198.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsieae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Madrid E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RA "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.";
RT Nature 396:133-140(1998).
RL Nature 396:133-140(1998).
CC -----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

```
CC -----
DR EMBL; AJ235270; CAA14663.1; -.
KW Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 355 HYPOTHETICAL PROTEIN RP198.
SQ SEQUENCE 355 AA; 40780 MW; A08B7655BEC86E0 CRC64;
```



Query Match 7.5%; Score 100; DB 1; Length 355;  
Best Local Similarity 25.7%; Pred. No. 0.85;  
Matches 66; Conservative 38; Mismatches 87; Indels 66; Gaps 16;

QY 2 NKPL-----LSLEIIDIGNTSIAFALEKDNQ-----VN--LFIKMKTNLMRYDEVIS 47  
Db 127 NKPLYKVSQDDSLKELKSIKLNKNTLSIFKDNQEFKINDLAFLIKKHNL-----178

QY 48 FFEENFDENVKVFISVVVPIILNETFKNVIFSFYKIKPLFIFGDLNYDLTFNPKSKDFL 107  
Db 179 -SQENISFLNMHYSE-KDILN-----FKNA-----NLDMATSFKAENGK-- 218

QY 108 LGSDFANLVAAIENYSF--ENVLVVDLGTACTIFAVSRQDGILGGIINSGPLINFSLL 165  
Db 219 -DAAILENL--NIERFICTDNESKVNLTGLOFFANKLPKGIL-----SFELENYNSIV 270

QY 166 D----NAYLIKKEPIST-PNNLLERTTSGSVNSGLFYQYKYLIEGVYRDIKQMYKKKFNL 220  
b 271 DKLLPNSILFSKTKTKTIIAKAMNKTSDQLNTD-----KNDTNSVYNNIK---NAKFDI 322

QY 221 IITGNADL-ILSLIEI 236  
Db 323 AFSDKGINIGSMNLEL 339

RESULT 8  
YC82\_EUGGR STANDARD; PRT; 516 AA.  
AC P30397;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical 64.3 kDa protein ycf82 (ORF516).  
GN YCF82.  
OS Euglena gracilis.  
OC Chloroplast.  
OC Eukaryota; Euglenozoa; Euglenida; Euglenales; Euglena.  
OX NCBI\_TaxID=3039;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-Z;  
RX MEDLINE-93347989; PubMed-8346031;  
RA Hallick R.B., Hong L., Drager R.G., Favreau M.R., Monfort A.,  
RA Orsat B., Spielmann A., Stutz E.;  
RT "Complete sequence of Euglena gracilis chloroplast DNA."  
RL Nucleic Acids Res. 21:3537-3544(1993).  
C -1- SIMILARITY: BELONGS TO THE YCF82 FAMILY. SOME SIMILARITY TO  
C GROUP II INTRON MATURASES.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).

-----  
DR EMBL; X70810; CAA50104.1; -.  
DR EMBL; Z11874; CAA77921.1; -.  
DR PIR; S26088; S26088.  
DR PIR; S34525; S34525.  
KW Chloroplast; Hypothetical protein.  
SQ SEQUENCE. 516 AA; 64333 MW; E2079EF48423490A CRC64;

Query Match 7.5%; Score 99.5; DB 1; Length 516;  
Best Local Similarity 20.0%; Pred. No. 1.4;  
Matches 42; Conservative 32; Mismatches 61; Indels 75; Gaps 8;

QY 87 FIGFDLN-----YDLTFNPKYKDKFLGSDVFANL-VAAIENYSFENVLVVDLGTACTIF 140  
Db 272 ILNININSFYRFETVFNFDFELRFGSNILLHINKKHIOFYKLSLKMIVKMFLLKKSUF 331

QY 141 AVSRQDGILGGIINSGPLINFSLLDNAYLIKKEPISTPNNLLERTTSGSVNSGLF---- 196  
Db 332 -----FLV-----NLLKKILDCLNRGFFCFNN 354

QY 197 -----YQYKYLIEGVYRDIKQMYKKK 217  
Db 355 NFLFLELDLYLRLL---WRYIKKLHSRK 380

RESULT 9  
YQIA\_CAEEL STANDARD; PRT; 971 AA.  
AC Q09281;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE Hypothetical 111.8 kDa protein C45G9.10 in chromosome III.  
GN C45G9.10.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-Bristol N2;  
RA Bentley D., Waterston R.;  
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).

-----  
DR EMBL; U21323; AAA62553.1; -.  
DR WormPep; C45G9.10; CE01851.  
KW Hypothetical protein.  
SQ SEQUENCE 971 AA; 111772 MW; 6AA5EB10E2BB7693 CRC64;

Query Match 7.4%; Score 99; DB 1; Length 971;  
Best Local Similarity 23.1%; Pred. No. 3.3;  
Matches 56; Conservative 37; Mismatches 83; Indels 66; Gaps 12;

QY 2 NKPLSELIIIDI-----GNTSIAFALEKDNQVNLFIKMKT-NLMLRYDEVYSFEEFNDFN 56  
Db 25 NNGLLTDKILKITSPHGNVSV---WDINEVIVFGEQESGKLSLILPKQFSIKLEKVDRT 80

QY 57 VNKVFISVVVPIILNETFKNVIFSFYKIKPLF--IGFDLN-----Y 94  
Db 81 FLNVFV--LKPIRDDYFHNVVIEYIVSKWEHWPSSLQFDLQSIHLKIVDDRSQSKDLILLY 138

QY 95 DLTFNPKYKDKFLLG---SDVFANLVAAIENYSFENVLV-----VDLGTACTI--- 139  
Db 139 DLRAEQQOTSKILLAPNKRDPKQCPAVMDTIHYSQVFQVQAMNRRSHVSSSRCTIAFL 198

QY 140 -----FAVSRQDGILGGIINSGPLINFSLLDNAYLIKKEPISTPN---NLLERT 186  
Db 199 LNTNGDFENYVTSKNDEI--RILNVTKIINAN---EYAYILVEFETLKPKGAENFVVQKQ 253

187 TS 188  
254 IS 255

RESULT 10  
Y478\_RICPR STANDARD; PRT; 554 AA.  
ID Y478\_RICPR

AC Q9ZD66;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein RP478 precursor.  
GN RP478.  
OS Rickettsia prowazekii.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsiae; Rickettsia.  
OX NCBI\_TaxID=782;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Madrid E;  
RX MEDLINE=99039499; PubMed=9823893;  
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,  
RA Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,  
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;  
RT "The genome sequence of Rickettsia prowazekii and the origin of  
RT mitochondria."  
RL Nature 396:133-140(1998).  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; AJ235271; CAA14933.1; --  
KW Hypothetical protein; Signal; Complete proteome.  
FT SIGNAL 1 33 POTENTIAL.  
FT CHAIN 34 554 HYPOTHETICAL PROTEIN RP478.  
SQ SEQUENCE 554 AA; 64341 MW; EC2C55EA9E2F373E CRC64;  
  
Query Match 7.2%; Score 96; DB 1; Length 554;  
Best Local Similarity 23.9%; Pred. No. 2.9;  
Matches 60; Conservative 28; Mismatches 81; Indels 82; Gaps 14;  
  
QY 3 KPLLSELIIDIGNTSIAFALFKDQVNLFIKMTNMLRYDEVYSFFEENFDNFVNKVF 62  
Db 264 KDIADKLTIKVNNAKINSNNF-ENNMLLYKGLN-----DFGNSNIHL 306  
QY 63 SSVVPILNETKKNVIFSFKKPLF-IGF---DINVDLTFNPKSKDFLLGSD--VFAN 115  
Db 307 S-----IESQFKLPGFIIGFLEFLKKNYD-----KQTYLLKFSDNKIYKN 347  
QY 116 LVAAIENYSFENVLVVDLTACTIFAVSRQDGILGGIINSGLINFNLSLDNAYLIKFFP 175  
348 -----FNNELVY-----ILNNKPNFNSILEDPRYHF-NFN 377  
QY 176 ISTPNLLERTTSGSVNSGLFYQ----YKYLIEGVYRDIKOMYKKKFNLIITGGNADLIL 231  
Db 378 INLVTE-LKKLTVQINTLSLYSNTSGFNITNETIINDLKDSYTKG---IIVINNSYKII 433  
QY 232 SLIEIEFIFNI 242  
Db 434 EILSF-YIYGV 443  
  
RESULT 11  
SYFB\_CAMJE STANDARD; PRT; 773 AA.  
ID SYFB\_CAMJE  
AC Q9PP35;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Phenylalanyl-tRNA synthetase beta chain (EC 6.1.1.20) (Phenylalanine--  
DE tRNA ligase beta chain) (PHERS).  
GN PHET OR CJ0896.  
OS Campylobacter jejuni.  
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;  
OC Campylobacter.

OX NCBI\_TaxID=197;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCTC 11168;  
RX MEDLINE=20150912; PubMed=10688204;  
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,  
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,  
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,  
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,  
RA Whitehead S., Barrell B.G.;  
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni  
RT reveals hypervariable sequences."  
RL Nature 403:665-668(2000).  
CC -!- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +  
CC diphosphate + L-phenylalanyl-tRNA(Phe).  
CC -!- SUBUNIT: Tetramer of two alpha and two beta chains (by  
CC SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- SIMILARITY: BELONGS TO THE PHENYLALANYL-TRNA SYNTHETASE BETA  
CC CHAIN FAMILY. SUBFAMILY 1.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; AL139076; CAB73154.1; --  
DR InterPro; IPR005146; B3\_4.  
DR InterPro; IPR005147; B5.  
DR InterPro; IPR005121; Fdx-AntiCB.  
DR InterPro; IPR003583; HHH\_1.  
DR InterPro; IPR004532; Phet\_bact.  
DR InterPro; IPR002547; tRNA\_bind.  
DR Pfam; PF01588; tRNA\_bind; 1.  
DR Pfam; PF03147; FDX-ACB; 1.  
DR Pfam; PF03483; B3\_4; 1.  
DR Pfam; PF03484; B5; 1.  
DR SMART; SM00278; Hhh1; 1.  
DR TIGRFAMS; TIGR00472; phet\_bact; 2.  
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
KW Complete proteome.  
SQ SEQUENCE 773 AA; 87250 MW; 21FE593C10AEFEF CRC64;  
  
Query Match 7.2%; Score 95.5; DB 1; Length 773;  
Best Local Similarity 22.3%; Pred. No. 4.7;  
Matches 48; Conservative 32; Mismatches 70; Indels 65; Gaps 9;  
  
QY 69 LNETFKNVIFSFKKIKPLFIFGDLNYDLTFNPKSKDFLLGSDVFANLVAAIEN----- 122  
Db 478 LNSAYKNYI-EFLNLR-----KRAVASGYFESLHYVLDNGEELKR 516  
QY 123 YSFENV-----LVVDLGTACTIFAVSRQDGILGGIINSGLINFNLSLDNAYLIKFFP 175  
Db 517 LGFDSVKLKLINPITAEIENLTTRT-----LLNHLNLAASLNAKNS----KKIILFE 564  
QY 176 ISTPNNL--LERTTSGSVNSGL-----FYQYKYLIEGVYRDIKOMYKKK 217  
Db 565 LGAVFNVNNQELNRIAFIHSGLKEAKISKAKPESVQVDFLLDIKNIIGDFK-LKSSK 623  
QY 218 FNLIITGGNADLILSLIEIEFIFNIHILTVGVRIL 252  
Db 624 YNILSPYEQADIYLSIDIKVGFGRHLHLKIENERDL 658  
  
RESULT 12  
SEC5\_YEAST  
ID SEC5\_YEAST STANDARD; PRT; 971 AA.  
AC P89102; Q04128; Q03775;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Exocyst complex component SEC5.  
GN SEC5 OR YDR166C OR YD9489.01C OR YD8358.20C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 69-80.  
RX MEDLINE=97133278; PubMed=8978675;  
RA Terbush D.R., Maurice T., Roth D., Novick P.;  
RT "The Exocyst is a multiprotein complex required for exocytosis in  
RL Saccharomyces cerevisiae.";  
RL EMBO J. 15:6483-6494(1996).  
RN [2]  
RP SEQUENCE OF 1-412 FROM N.A.  
RC STRAIN=S288c / AB972;  
RA Oliver K., Harris D., Barrell B.G., Rajandream M.A.;  
RA Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 81-971 FROM N.A.  
RC STRAIN=S288c / AB972;  
RA Murphy L., Richards C., Harris D., Barrell B.G., Rajandream M.A.,  
RA Walsh S.V.;  
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: REQUIRED FOR EXOCYTOSIS.  
CC -!- SUBUNIT: SEC3, SEC5, SEC6, SEC8, SEC10, SEC15 AND EXO70 ARE  
CC COMPONENT OF EXOCYST COMPLEX.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Y08788; CAA70040.1; -.  
DR EMBL; 247813; CAA87797.1; -.  
DR EMBL; Z50046; CAA90386.1; -.  
DR SGD; S0002573; SEC5.  
KW Transport; Protein transport; Golgi stack; Coiled coil.  
FT DOMAIN 123 182 COILED COIL (POTENTIAL).  
SQ SEQUENCE 971 AA; 112121 MW; E41AC5A15CE3D1DF CRC64;  
Query Match 7.1%; Score 95; DB 1; Length 971;  
Best Local Similarity 21.6%; Pred. No. 6.8;  
Matches 64; Conservative 41; Mismatches 115; Indels 76; Gaps 13;  
Qy 6 LSELIIDIGNTSIAFALFKDQVNLFIKMKTNLMRLRYDEVYSFFFEENFDF-----NVN 58  
Db 133 LDRLDSIDIQDSIHLKQLVGKFTKYVKKIKNKL---DQYKEFDEKTEKNQCDSPEKN 188  
Qy 59 KVFISVVPILNETFKNVI-FSFFKIKPLFIGFD--LNYDLT-----FNPKSKDK 105  
Db 189 QINVES---LNKKVDEVIRFTTFKLPMDNYQKILNYQATKKFIELNKFYFNLPSKSLK 244  
Qy 106 FLGSDVVFANLVAAIENYS-----FENVLVLDGTACTIFAVSRQDGIILG 150  
Db 245 RCLTNNDNFNFI--EYSKGLTLRRRNFNQSSDASQSLVKKRIWTQIENLLVTYKDLIWN 301  
Qy 151 GIINSGPLINFN-----SLLDNAYLIKFKPISTPNLLE-----RTTSGSVNSGL 195  
Db 302 SLINS---NFNIDQPQETILSLFSKLLNLENF---INNORESESGNKNTTSSSNENPI 354  
Qy 196 FYQKYLLIEGVYRDIKQYKKKFNLI-----TGGNADLILSLIEFIFNIH 243  
Db 355 LRWMSIKMNGFQNELNELSGHMSKIIHSORLILQNNTNODKSGCGVELSYLKN 410  
RESULT 13  
MVIN\_RICPR  
ID MVIN\_RICPR STANDARD; PRT; 507 AA.

AC Q9ZCW4;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Virulence factor mvIN homolog.  
GN MVIN OR RP590.  
OS Rickettsia prowazekii.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsieae; Rickettsia.  
OX NCBI\_TaxID=782;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Madrid E;  
RX MEDLINE=99039499; PubMed=9823893;  
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,  
RA Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,  
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;  
RT "The genome sequence of Rickettsia prowazekii and the origin of  
RL mitochondria.";  
RL Nature 396:133-140(1998).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -!- SIMILARITY: BELONGS TO THE MVIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AJ235272; CAA15035.1; -.  
DR InterPro; IPR004268; MVIN\_like.  
DR Pfam; PF03023; MVIN; 1.  
KW Transmembrane; Complete proteome.  
FT TRANSMEM 3 23 POTENTIAL.  
FT TRANSMEM 54 74 POTENTIAL.  
FT TRANSMEM 92 112 POTENTIAL.  
FT TRANSMEM 132 152 POTENTIAL.  
FT TRANSMEM 157 177 POTENTIAL.  
FT TRANSMEM 185 205 POTENTIAL.  
FT TRANSMEM 245 265 POTENTIAL.  
FT TRANSMEM 268 288 POTENTIAL.  
FT TRANSMEM 310 330 POTENTIAL.  
FT TRANSMEM 351 371 POTENTIAL.  
FT TRANSMEM 379 399 POTENTIAL.  
FT TRANSMEM 405 425 POTENTIAL.  
FT TRANSMEM 438 458 POTENTIAL.  
FT TRANSMEM 472 492 POTENTIAL.  
SQ SEQUENCE 507 AA; 56059 MW; FFC56CCE2EC8EAC1 CRC64;  
Query Match 7.0%; Score 93.5; DB 1; Length 507;  
Best Local Similarity 22.0%; Pred. No. 4.1;  
Matches 63; Conservative 58; Mismatches 112; Indels 53; Gaps 14;  
Qy 2 NKPLLSELIIDIGNTSIA-----FALFKDQVNLFIKMKTNLMRLRYDEVYSFFFEENFDFN 56  
Db 220 NDPDVKKLLINMGPATISSGVQQLNLFISQSISSFIEGAISILAYADRIYQFPLS----I 275  
Qy 57 VNKVFISVVPILNETFK-NVIFSFFKIK--PLFIGFDLNYDLTFNPKSKDKFLGSDVF 113  
Db 275 IGTSFSTILPEMSKVYKSNDIVSAQKIQNNNAIRIGLLSLPATEFG-----IILSHPI 329  
Qy 114 ANLVAAIENYSFENVLVLDGTACTIFAVSRQDGIILGGIINSGPLINFNLSLLDNAYLIK 173  
Db 330 TNII--YERGVTTPQDTTNTAEAISAFALGPLAFILAKILT--PIFYANGDTKTKPLKITL 385  
Qy 174 FP--ISTPNLL-----ERTTSGSVNSGLFYQYK-----YLIEGVYRDIKQM 213  
Db 386 FSIINTNMNLLMLDSLKHIGIAGVGTSAAWYNLGLLYSYSTKQHKHIEAGI-----KL 440  
Qy 214 YKKKFNLIITGGNADLILSLIE---IEFIFNIHLTVEGVRIILGNSI 256



Db	441	PCAKILLCCT--LMSIIIALIKHYLYSEYLLIK-VSMLGSTI	483
Db	14	RESULT 14	
Y468_MYCGE			
ID	Y468_MYCGE	STANDARD;	PRT; 1783 AA.
AC	Q49460;		
DT	01-NOV-1997	(Rel. 35, Created)	
DT	15-JUL-1999	(Rel. 38, Last sequence update)	
DT	16-OCT-2001	(Rel. 40, Last annotation update)	
DE	Hypothetical protein MG468.		
GN	MG468.		
OS	Mycoplasma genitalium.		
OC	Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.		
OX	NCBI_TaxID=2097;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=ATCC 33530 / G-37;		
RX	MEDLINE=96026346; PubMed=7569993;		
RA	Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A., Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M., Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L., Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M., Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S., Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.; "The minimal gene complement of Mycoplasma genitalium.";		
RL	Science 270:397-403(1995).		
RN	[2]		
RP	REVISIONS.		
RA	Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A., Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M., Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L., Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M., Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S., Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.; Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.		
RN	[3]		
RP	SEQUENCE OF 879-985 FROM N.A.		
RC	STRAIN=ATCC 33530 / G-37;		
RX	MEDLINE=94075230; PubMed=8253680;		
RA	Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;		
RT	"A survey of the Mycoplasma genitalium genome by using random sequencing.";		
RL	J. Bacteriol. 175:7918-7930(1993).		
CC	-!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).		
CC	-!- SIMILARITY: SOME, TO MG064.		
CC			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC			
DR	EMBL; U39728; AAC72488.1; -		
DR	EMBL; U01808; AAD12339.1; -		
DR	TIGR; MG468; -		
KW	Hypothetical protein; Transmembrane; Complete proteome.		
FT	TRANSMEM 16 36 POTENTIAL.		
FT	TRANSMEM 917 937 POTENTIAL.		
FT	TRANSMEM 967 987 POTENTIAL.		
FT	TRANSMEM 1010 1030 POTENTIAL.		
FT	TRANSMEM 1084 1104 POTENTIAL.		
FT	TRANSMEM 1660 1680 POTENTIAL.		
FT	TRANSMEM 1709 1729 POTENTIAL.		
FT	TRANSMEM 1730 1750 POTENTIAL.		
FT	TRANSMEM 1752 1772 POTENTIAL.		
SEQUENCE	1783 AA; 200168 MW; 87BD575AEC2E374B CRC64;		
Query Match			
Best Local Similarity	7.0%; Score 93.5; DB 1; Length 1783;		
Matches	60; Conservative 36; Mismatches 81; Indels 95; Gaps 12;		
QY	17	SIAFALFKDNQVNLFIKMKTNLMRLRYDE-----VYSFFEENFD-----FNVNKV-----	60
Db	838	SILFAVPAANQENYYAFKSTDLKQHTDQDPVQFIANRLEGYLDVPRSDLAFAFNVDISKFN	897
QY	61	-----FISSVVPILNETKFNVIFFSKIKPLFIFIGEDLNVDLTFN-----	99
Db	898	YLTARNYFPDLVQSYLAIVSTVIAIF-----LIILALYLIILLIKSEFIKKNQTEFSIIR	952
QY	100	--PYKSDKFLLGSDVFANLVAAIENYSFENVLVVDLGTACTIFAVSRQDGLGGIISGP	157
Db	953	AGGFSTTKFIVGMSVFAGIVAIVS--SFLGLVLAFL-----LEGQVKGIIN----	996
QY	158	LINFNSLLDNAYLIKKFPISPTPNLLERTTSGSVN---SGLFYQYKYLIIEGV-YRDIKQ	212
Db	997	-----RYWFIALPEN-----SFNWLSTFGSFFITFFVFEFISWIAFKQ	1034
QY	213	MYKKKFNLIITGGNADLILSLIEIEFIFNIHL	244
Db	1035	LFSKPVNVLIDQGN-----ETKFSVLLHL	1058
RESULT 15			
ANFK_AZOVI			
ID	ANFK_AZOVI	STANDARD;	PRT; 462 AA.
AC	P16267;		
DT	01-APR-1990	(Rel. 14, Created)	
DT	01-APR-1990	(Rel. 14, Last sequence update)	
DT	15-JUL-1999	(Rel. 38, Last annotation update)	
DE	Nitrogenase iron-iron protein beta chain (EC 1.18.6.1) (Nitrogenase component I) (Dinitrogenase 3 beta subunit).		
GN	ANFK.		
OS	Azotobacter vinelandii.		
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;		
OC	Azotobacter.		
OX	NCBI_TaxID=354;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=89123105; PubMed=2644222;		
RA	Joergers R.D., Jacobson M.R., Premakumar R., Wolfinger E.D., Bishop P.E.;		
RT	"Nucleotide sequence and mutational analysis of the structural genes (anfHDKG) for the second alternative nitrogenase from Azotobacter vinelandii.";		
RL	J. Bacteriol. 171:1075-1086(1989).		
CC	-!- FUNCTION: THE KEY ENZYMAIC REACTIONS IN NITROGEN FIXATION ARE CATALYZED BY THE NITROGENASE COMPLEX, WHICH HAS 2 COMPONENTS: THE IRON PROTEIN (COMPONENT 2) AND A COMPONENT 1 WHICH IS EITHER A MOLYBDENUM-IRON PROTEIN, A VANADIUM-IRON, OR AN IRON-IRON PROTEIN.		
CC	-!- CATALYTIC ACTIVITY: 8 reduced ferredoxin + 8 H(+) + N(2) + 16 ATP -> 8 oxidized ferredoxin + 2 NH(3) + 16 ADP + 16 phosphate.		
CC	-!- COFACTOR: IRON-SULFUR.		
CC	-!- SUBUNIT: HEXAMER OF TWO ALPHA, TWO BETA, AND TWO DELTA CHAINS.		
CC	-!- SIMILARITY: BELONGS TO THE NIFD/NIFK/NIFE/NIFN FAMILY.		
CC			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC			
DR	EMBL; M23528; AAA82511.1; -		
DR	PIR; D32057; D32057.		
DR	HSSP; P11347; LMIO.		
DR	InterPro; IPR000318; Nitrognase_compl.		
DR	InterPro; IPR000510; Oxred_nitrognasel.		
DR	Pfam; PF00148; oxidored_nitro; 1.		
DR	PROSITE; PS00090; NITROGENASE_1_2; 1.		
DR	PROSITE; PS00699; NITROGENASE_1_1; 1.		
KW	Oxidoreductase; Nitrogen fixation; Iron-sulfur.		
SEQUENCE	462 AA; 51179 MW; 2E4DE6267094E3CC CRC64;		



Query Match 7.0%; Score 93; DB 1; Length 462;  
Best Local Similarity 21.7%; Pred. NO. 4;  
Matches 65; Conservative 44; Mismatches 88; Indels 102; Gaps 15;

QY 1 MNKPLLSELIIDIGNTSIAFALFKDNQVNLFIKMKTNMLRLRYDEVYSFEEFENFENVKV 60  
Db 120 LNEGLLKE-----KFPDREVHL-IAHMT-----PS 143

QY 61 FISSVVPILNETFKNVIFSEFKIKP-----LFIGFDLNYDLTFNP--YKSDKFLLGS-D 111  
Db 144 FVGS MISGYDVAVRDVVRHFHFAKREAPNDKINLLTGW-----VNP GDVKELKHLGEMD 196

QY 112 VFANLVAAIENY-----SFENVLV---VDLGTACTIFAVSRQDGILGGIINSGP 157  
Db 197 IEANVLFEIESFSDSPILPDGSAVSHGNTTIEDLIDTGNARATFALNRYEGTKAA----- 250

QY 158 LINFNSLLDNAYLIKPPPI-----STPNNL-----LERTTSGSVNSGLFYQYKYL 203  
b 251 -----EYLQKKFEIPAIIGPTPIGIRNTDIFLQNLKKATGKPIQPSLAHERGVAI 300

QY 204 EGYRDIKQMYKKKFNLIITGKNADLILSLIEIEIFNHLTVGVRIIGNSIDFKFVN 262  
Db 301 DAL-ADLTHMFLAEKRVAIYGA-PDLVIGLAE---FCLDLEMKPVLLLLGDDNSKYVD 353

Search completed: June 24, 2003, 21:51:19  
Job time : 6.55691 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:36:04 ; Search time 9.55623 Seconds  
(without alignments)  
2635.685 Million cell updates/sec

Title: US-09-813-453A-11  
Perfect score: 1331  
Sequence: 1 MNKPLLSELIIDIGNTSIAF.....HLTVEGVRILGNSIDFKFVN 262

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

---Searched: 283224 seqs, 96134422 residues

---Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1331	100.0	262	2 F70165	conserved hypothet
2	289.5	21.8	273	2 E97293	probable transcrip
3	275.5	20.7	254	2 F83660	hypothetical prote
4	253.5	19.0	259	2 AF1464	conserved hypothet
5	248.5	18.7	259	2 AF1102	conserved hypothet
6	231.5	17.4	233	2 S66100	conserved hypothet
7	203	15.3	246	2 D72320	conserved hypothet
8	189.5	14.2	261	2 B87489	transcription acti
9	181.5	13.6	265	2 T36391	hypothetical prote
10	170	12.8	273	2 D71326	conserved hypothet
11	159	11.9	262	2 E75516	conserved hypothet
12	152.5	11.5	224	2 A99571	conserved hypothet
13	143	10.7	274	2 H86937	conserved hypothet
14	141.5	10.6	229	2 E70465	hypothetical prote
15	138	10.4	272	2 A70955	hypothetical prote
16	121.5	9.1	209	2 H81382	hypothetical prote
17	120.5	9.1	476	2 A28439	endonuclease SceI
18	118	8.9	276	2 AI2292	hypothetical prote
19	118	8.9	625	2 A90127	hypothetical prote
20	116	8.7	2178	2 S55805	alpha-toxin - Clos
21	112.5	8.5	963	2 C90535	conserved hypothet
22	112	8.4	636	2 F90094	hypothetical prote
23	111.5	8.4	629	2 T28217	hypothetical prote
24	110.5	8.3	1003	2 H82883	hypothetical prote
25	110	8.3	300	1 S41759	ribosomal protein
26	109.5	8.2	2136	2 A05037	hypothetical prote
27	108	8.1	311	2 D97058	SAM-dependent meth
28	107.5	8.1	1613	2 D90129	hypothetical prote
29	107	8.0	590	2 G90127	hypothetical prote

30	103.5	7.8	1581	2 B71636	hypothetical prote
31	103	7.7	436	2 B70157	hypothetical prote
32	103	7.7	1465	2 A70199	hypothetical prote
33	102	7.7	1024	1 RNZQBF	DNA-directed RNA p
34	101.5	7.6	594	2 A82913	hypothetical prote
35	101.5	7.6	1119	2 T18491	hypothetical prote
36	101.5	7.6	1802	2 G71616	hypothetical prote
37	101	7.6	1175	2 F64489	hypothetical prote
38	100.5	7.6	495	2 E70198	hypothetical prote
39	100.5	7.6	512	2 F71915	hypothetical prote
40	100.5	7.6	568	2 D90525	hypothetical prote
41	100	7.5	355	2 H71730	hypothetical prote
42	100	7.5	592	2 B81009	hypothetical prote
43	99.5	7.5	410	2 S72277	BirA protein/Bvg a
44	99.5	7.5	439	2 T28196	translation elonga
45	99.5	7.5	516	2 S34525	hypothetical prote

ALIGNMENTS

RESULT 1

F70165

conserved hypothetical protein BB0527 - Lyme disease spirochete

C;Species: Borrelia burgdorferi (Lyme disease spirochete)

C;Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 08-Oct-1999

C;Accession: F70165

R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; W son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; ; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, E Nature 390, 580-586, 1997

A;Authors: Smith, H.O.; Venter, J.C.

A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.

A;Reference number: A70100; MUID:98065943; PMID:9403685

A;Accession: F70165

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-262 <KLE>

A;Cross-references: GB:AE001154; GB:AE000783; NID:g2688431; PIDN:AAC66882.1; PID:g26

A;Experimental source: strain B31

Query Match 100.0%; Score 1331; DB 2; Length 262;  
Best Local Similarity 100.0%; Pred. No. 5.6e-94;  
Matches 262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MNKPLLSELIIDIGNTSIAFALFKDNQVNLFIKMTNLMRLRYDEVYSFFEENFDNFVKV	60
DB	1		
QY	61	FISSVVPILNETFKNVIFSFKKIKPLFIGFDLNYDLTFNPKSKDKFLGSDVFANLVA	120
DB	61	FISSVVPILNETFKNVIFSFKKIKPLFIGFDLNYDLTFNPKSKDKFLGSDVFANLVA	120
QY	121	ENYSFENVLVDLGTACTIFAVSRQDGIILGGIINSGLPLNENSLLDNAYLIKKFPISTPN	180
DB	121	ENYSFENVLVDLGTACTIFAVSRQDGIILGGIINSGLPLNENSLLDNAYLIKKFPISTPN	180
QY	181	NLLERTTSGSVNSGLFYQYKYLIEGVYRDIKQYKKKFNLIITGGNADLILSLIEIEFIF	240
DB	181	NLLERTTSGSVNSGLFYQYKYLIEGVYRDIKQYKKKFNLIITGGNADLILSLIEIEFIF	240
QY	241	NIHLTVEGVRILGNSIDFKFVN 262	
DB	241	NIHLTVEGVRILGNSIDFKFVN 262	

RESULT 2

E97293

probable transcription regulator, homolog of Bvg accessory factor [imported] - Clost

C;Species: Clostridium acetobutylicum

C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001

C;Accession: E97293

R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.;







```
RESULT 8
B87489
transcription activator, probable Baf family [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C;Accession: B87489
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n.; J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: B87489
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-261 <STO>
A;Cross-references: GB:AE005673; NID:gl3423392; PIDN:AAK23910.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC1935
;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c
Query Match 14.2%; Score 189.5; DB 2; Length 261;
Best Local Similarity 26.9%; Pred. No. 2.8e-07;
Matches 67; Conservative 41; Mismatches 130; Indels 11; Gaps 4;
QY 9 LIIDIGNTSIAFALKDNQVNLFTKMKTNLMRLRYDEVYSFFEENFDN-----VNKVF 62
Db 1 LAIEQNTNTMFAIHGDGASWAQWRSATESRTRDAEYVVWLSQLLSMQGLGFRAIDAVII 63
QY 63 SSVVPILNETFKNVIFSFFKIKPLFGFDLNYDTFNPKYKSDKFLGSDVFANLVA 122
Db 64 SSVVPOSIFNLRLNRRYFNEPLVIGENAKLGIDVRIKPS--AGADRLVNAIGA 121
QY 123 YSFENVLVVDLGTACTIFAVSRQDGLIGLINSGLPLNFNSLLDNAYLIKKEPISTP--N 180
Db 122 YP-GPLVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMQALHEAAKLPRIAIORPAGN 180
QY 181 NLLERTSGSVNSGLFYQYKYLIEGVYRDIKOMYKKKFNLIITGGNADLILSLIEIEFIF 240
Db 181 RIVGTDVTSAMQSGVFWGYISLIEGLVARIKAEERGEPTMTVIATGGVASFEGATDSIDHF 240
QY 241 NIHLTVEGV 249
Db 241 DSDLTIRGL 249
RESULT 9
T36391
'hpothetical protein SCE94.31c - Streptomyces coelicolor
;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jun-2000
C;Accession: T36391
R;Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, April 1999
A;Reference number: Z21573
A;Accession: T36391
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-265 <OLI>
A;Cross-references: EMBL:AL049628; PIDN:CAB40880.1; GSPDB:GN00070; SCOEDB:SCE94.31c
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SCE94.31c
;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c
Query Match 13.6%; Score 181.5; DB 2; Length 265;
Best Local Similarity 23.6%; Pred. No. 1.2e-06;
Matches 63; Conservative 55; Mismatches 114; Indels 35; Gaps 7;
QY 9 LIIDIGNTSIAFALKDNQVNLFIKMKTNLMRLRYDE-----VYSFFEENFDNFVN 58
Db 3 LTIDVGNTHTVLGLFDGEDIVEHWRISTDSRRRTADELAVLLQGLMGHPLLGLDELGDGID 62
```

```
QY 59 KVFISVVPIILNETFKNVIFSFFKIKPLFI-----GFDLNYDLTFNPKYKSDKFLLGSD 111
Db 63 GIAICATVPSVLHELREVTTRYGYDPAVLVEPGVKTGVP-----LTDHPKE-----VGAD 114
QY 112 VFANLVAIAIENYSFENVLVVDLGTACTIFAVSRQDGLIGLINSGLPLNFNSLLDNAYLI 171
Db 115 RIINAVAAVELYG-GPAIVVDFTATTFDVARSARGEIVGGIAPGIEISVEALGVKGAQL 173
QY 172 KKFPISIPNNLLERTSGSVNSGLFYQYKYLIEGVYRDIKOMYKK-----KFNLIITGG 225
Db 174 RKIEVARPRSVIGKNTVEAMQSGIVYGFAGQVDGV---VNRMARELADDPDDVTVIATGG 230
QY 226 NADLILSLIEIEFIFNIHLTVEGVRL 252
Db 231 LAPMVLGESSVIDHEHPWLTLMGLRLV 257
RESULT 10
D71326
conserved hypothetical protein TP0431 - syphilis spirochete
C;Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C;Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 20-Jun-2000
C;Accession: D71326
R;Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.;
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.;
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A;Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A;Reference number: A71250; MUID:98332770; PMID:9665876
A;Accession: D71326
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-273 <COL>
A;Cross-references: GB:AE001220; GB:AE000520; NID:g3322705; PIDN:AAC65417.1; PID:g33:
A;Experimental source: strain Nichols
C;Genetics:
A;Gene: TP0431
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c
Query Match 12.8%; Score 170; DB 2; Length 273;
Best Local Similarity 25.3%; Pred. No. 9e-06;
Matches 64; Conservative 50; Mismatches 123; Indels 16; Gaps 6;
QY 9 LIIDIGNTSIAFALKDNQ-----VNLFIKMKTNLMRLRYDEVYSFFEENF-----DFNV 57
Db 2 LLIDVGNSHVVFQIGENGGRVCVRELFRLAPDARKTQDE-YSLLIHALCERAGVGRASL 60
QY 58 NKVFISVVPIILNETFKNVIFSFFKIKPLFGFDLNYDTFNPKYKSDKFLGSDVFANLV 117
Db 61 RDAFISVVPIILNETFKNVIFSFFKIKPLFGFDLNYDTFNPKYKSDKFLGSDVFANLV 120
QY 118 AAIENYSFENVLVVDLGTACTIFAVSRQDGLIGLINSGLPLNFNSLLDNAYLIKKEPIS 177
Db 121 AAYVHFR-SACVVVDGCGTALTFTTAVDGTGLIQGVAIAPGLRTAVQSLHTGTATLPLVPLA 179
QY 178 TPNNLLERTSGSVNSGLFYQYKYLIEGVYRDIKOMYKKKFNLIITGGNADLILSLIEIE 237
Db 180 LPDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCQKELGCRCAAVITGGLSRLFSS--EVD 237
QY 238 F-IFNIHLTVEGV 249
Db 238 FPPIDAQLTSLGL 250
RESULT 11
E75516
conserved hypothetical protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 20-Jun-2000
C;Accession: E75516
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
```







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:26 ; Search time 26.899 Seconds  
(without alignments)  
1297.879 Million cell updates/sec

Title: US-09-813-453A-11  
Perfect score: 1331  
Sequence: 1 MNKPLLSLELIIDIGNTSIAF.....HLTVEGVRILGNSIDPKFVN 262

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues  
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1331	100.0	262	AAU91158	Borrelia burgdorferi
2	282	21.2	256	AAU91175	Pantothenate kinas
3	276.5	20.8	262	AAU91170	Pantothenate kinas
4	275.5	20.7	254	AAU91171	Pantothenate kinas
5	271.5	20.4	258	AAU01243	B. subtilis novel
6	271.5	20.4	258	AAU91149	Bacillus subtilis
7	266.5	20.0	255	AAU91154	Geobacter sulfurre
8	262	19.7	250	AAU91150	Clostridium acetob
9	256.5	19.3	258	AAU91172	Pantothenate kinas
10	248.5	18.7	259	ABB47661	Listeria monocytog

11	231.5	17.4	233	23	AAU91163	Pantothenate kinas
12	203	15.3	246	23	AAU91156	Thermotoga maritim
13	197.5	14.8	258	23	AAU91153	Rhodobacter capsul
14	190	14.3	219	23	AAU91176	Pantothenate kinas
15	189.5	14.2	260	23	AAU91173	Pantothenate kinas
16	181.5	13.6	265	23	AAU91151	Streptomyces coeli
17	170	12.8	273	23	AAU91157	Treponema pallidum
18	166.5	12.5	257	23	AAU91174	Pantothenate kinas
19	159	11.9	262	23	AAU91155	Deinococcus radiop
20	141.5	10.6	229	23	AAU91159	Aquifex aeolicus p
21	138	10.4	272	22	AAG81225	Mycobacterium tube
22	138	10.4	272	23	AAU91152	Mycobacterium tube
23	136	10.2	241	23	AAU91179	Pantothenate kinas
24	135	10.1	244	23	AAU91168	Pantothenate kinas
25	121.5	9.1	209	23	AAU91165	Pantothenate kinas
26	121.5	9.1	476	21	AAU54039	Amino acid sequenc
27	121.5	9.1	503	13	AAR22667	50 kD subunit of S
28	111	8.3	212	23	AAU91177	Pantothenate kinas
29	104	7.8	471	21	AAU54040	Pantothenate kinas
30	101.5	7.6	1802	21	AAB18217	Amino acid sequenc
31	100	7.5	389	21	AAU74909	Plasmodium falcipa
32	100	7.5	592	20	AAU38615	Neisseria meningit
33	100	7.5	592	21	AAU74912	Neisseria meningit
34	100	7.5	592	23	AAU91169	Neisseria meningit
35	99.5	7.5	410	22	AAE07211	Pantothenate kinas
36	99.5	7.5	410	23	AAU98959	Plasmodium falcipa
37	98.5	7.4	468	23	ABB77623	Plasmodium falcipa
38	97.5	7.3	1188	21	AAB18183	AmEPV serine threo
39	96.5	7.3	439	23	ABP26082	Plasmodium falcipa
40	96	7.2	455	21	AAU74910	Streptococcus poly
41	96	7.2	592	20	AAU38616	Neisseria meningit
42	96	7.2	592	21	AAU74913	Neisseria meningit
43	96	7.2	592	23	AAU91166	Neisseria meningit
44	95.5	7.2	810	23	ABP35711	Pantothenate kinas
45	95	7.1	451	21	AAU59283	Fungal ZBC protein
						WAR toxin from B.

ALIGNMENTS

RESULT 1  
AAU91158  
ID AAU91158 standard; Protein; 262 AA.  
XX  
AC AAU91158;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Borrelia burgdorferi pantothenate kinase Coax.  
XX  
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX  
OS Borrelia burgdorferi.  
XX  
PN WO200216601-A2.  
XX  
PD 28-FEB-2002.  
XX  
PF 24-AUG-2001; 2001WO-US26531.  
XX  
PR 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX  
(OMNI-) OMNIGENE BIOPRODUCTS INC.  
PI Yocum RR, Patterson TA;  
XX  
DR WPI; 2002-269358/31.  
XX  
PT Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein

XX Claim 6; Page 76; 128pp; English.  
PS The invention describes assays for identifying a (potential) antibiotic  
XX comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
XX described in the invention.  
SQ Sequence 262 AA;

Query Match 100.0%; Score 1331; DB 23; Length 262;  
Best Local Similarity 100.0%; Pred. No. 3.1e-127;  
Matches 262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 MNKPLLSLLIDIGNTSIAFALFKDQVNLFIKMKTNLMRLRYDEVYSFFFEENFDFNVKV 60  
1 MNKPLLSLLIDIGNTSIAFALFKDQVNLFIKMKTNLMRLRYDEVYSFFFEENFDFNVKV 60  
61 FISSVVPILNETFKNVIFSEFFKIKPLFIQFDLNYDLTFNPKSKDKFLLGSDVFANLVA 120  
61 FISSVVPILNETFKNVIFSEFFKIKPLFIQFDLNYDLTFNPKSKDKFLLGSDVFANLVA 120  
121 ENYSFENVLVVDLGTACTIFAVSRQDGLGGIINSGPLINFNLSLLDNAYLIKKEPISTPN 180  
121 ENYSFENVLVVDLGTACTIFAVSRQDGLGGIINSGPLINFNLSLLDNAYLIKKEPISTPN 180  
181 NLLERTSGSVNSGLFYQYKYLIEGVYRDIKQYKKKFNLIITGGNADLILSLIEIEFIF 240  
181 NLLERTSGSVNSGLFYQYKYLIEGVYRDIKQYKKKFNLIITGGNADLILSLIEIEFIF 240  
241 NIHLTVGVRILGNSIDFKFVN 262  
241 NIHLTVGVRILGNSIDFKFVN 262

RESULT 2  
AAU911175  
ID AAU911175 standard; Protein; 256 AA.  
XX AAU911175;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
Pantothenate kinase (Coax) #13.  
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX Clostridium difficile.  
OS WO200216601-A2.  
XX  
PN 28-FEB-2002.  
XX  
PD 24-AUG-2001; 2001WO-US26531.  
XX  
PF 24-AUG-2000; 2000US-227860P.  
XX  
PR 20-MAR-2001; 2001US-0813453.  
XX  
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX  
XX Yocum RR, Patterson TA;  
PI WPI; 2002-269358/31.  
XX  
DR N-PSDB; ABK54196.  
XX  
PT Identifying potential antibiotic or antimicrobial agent, comprises

PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -  
XX Claim 6; Page 105; 128pp; English.  
PS The invention describes assays for identifying a (potential) antibiotic  
XX comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
XX described in the invention.  
SQ Sequence 256 AA;

Query Match 21.2%; Score 282; DB 23; Length 256;  
Best Local Similarity 31.3%; Pred. No. 2e-20;  
Matches 82; Conservative 54; Mismatches 94; Indels 32; Gaps 10;  
9 LIIDIGNTSIAFALFKDQVNLFIKMKTNLMRLRYDEVYSFFFEENFDF-NVN-----KVFI 62  
3 LVFDVGTNMVLGIYKDKLVNWRIKTDREKTSDEYGILISNLFYDYNVISDIDDDVII 62  
63 SSVVPILNETFKNVIFSEFFKIKPLFI-----GFDLNYDLTFNPKSKDKFLLGSDVFANL 116  
63 SSVVPNMVHLSLENFCIKYCKKQPLIVGPGIKTGLNIKYD---NPKQ-----VGADRIVNA 114  
117 VAAIENYSFENVLVVDLGTACTIFAVSRQDGLGGIINSGPLINFNLSLLDNAYLIKKEPI 176  
115 VAGIEKYGAPSIL-VDFGTATTFCAISEKGEYLGGTIAPGKISSEALFQSASKLPRVEL 173  
177 STPNLLERTSGSVNSGLFYQYKYLIEGVYRDIKQYKKKFN-----LIITGGNADLIL 231  
174 AKPGMTICKSTVSAMQSGIYGVGLVD---KIISMKKELNCDDVKVIATGGLAKLIA 229  
232 SLIE-IEFIFNIHLTVGVRIL 252  
230 SETKSIDYVDGF-LTLEGLRII 250

RESULT 3  
AAU911170  
ID AAU911170 standard; Protein; 262 AA.  
XX AAU911170;  
AC  
XX 05-JUN-2002 (first entry)  
DT  
XX Pantothenate kinase (Coax) #8.  
DE Pantothenate kinase; Coax; antibiotic; antimicrobial;  
XX pantothenate kinase modulator; coenzyme A; bactericidal compound.  
KW Bacillus anthracis.  
XX WO200216601-A2.  
XX  
PN 28-FEB-2002.  
XX  
PD 24-AUG-2001; 2001WO-US26531.  
XX  
PF 24-AUG-2000; 2000US-227860P.  
XX  
PR 20-MAR-2001; 2001US-0813453.  
XX  
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX  
XX Yocum RR, Patterson TA;  
PI WPI; 2002-269358/31.  
XX  
DR N-PSDB; ABK54191.

XX Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -  
XX  
XX Claim 8; Page 98-99; 128pp; English.  
CC The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.  
XX  
XX Sequence 262 AA;  
Query Match 20.8%; Score 276.5; DB 23; Length 262;  
Best Local Similarity 27.8%; Pred. No. 7.6e-20;  
Matches 71; Conservative 63; Mismatches 100; Indels 21; Gaps 7;  
QY 10 IIDIGNTSIAFALFKDNQVNLFIKMTNMLRYDE----VYSFFE-ENFDF-NVNKVFIS 63  
Db 4 VLDVGNTNAVLGVEEGELRQHRMETDRHKTEDEYGMVLVKQLLEHGLSFEDVKGIIVS 63  
QY 64 SVVPILNETKKNVIFSFFKIKPLFI-----GFDLNYDLTFNPKYKSDKFLGSDVFANLV 117  
Db 64 SVVPPIMFALERMCEKYPKIKPLVWGPGIKTGLNIKYE--NPKE-----VGADRIVNAV 115  
QY 118 AAIENYSFENVLVLDGTACTIFAVSRQDGILGGIINSGPLINFNLSLDNAYLIKKEPIS 177  
Db 116 AGIHLYG-SPLIIVDFGTATTCYIINEEKHYMGVITPGIMISAEALYSRAAKLPRIET 174  
QY 178 TPNNLLERTTSGSVNSGLFYQYKYLIEGVYRDIKQYKKKFNLIITGGNADLILSLIEIE 237  
Db 175 KPSSVVGKNTVSAMQSGILYGVQVEGIVKRMKEAKQEPKVIATGGLAKLISESNVI 234  
QY 238 FIFNIHLTVEGVRIL 252  
Db 235 DVVDPFLTLKGLYML 249  
RESULT 4  
AAU91171 standard; Protein; 254 AA.  
X  
AC AAU91171;  
XX  
DT 05-JUN-2002 (first entry)  
XX Pantothenate kinase (Coax) #9.  
DE Pantothenate kinase; Coax; antibiotic; antimicrobial;  
XX Pantothenate kinase modulator; coenzyme A; bactericidal compound.  
KW Bacillus halodurans.  
XX OS WO200216601-A2.  
XX PN 28-FEB-2002.  
XX PD 24-AUG-2001; 2001WO-US26531.  
XX PF 24-AUG-2000; 2000US-227860P.  
XX PR 20-MAR-2001; 2001US-0813453.  
XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX PI Yocum RR, Patterson TA;  
XX

DR WPI; 2002-269358/31.  
DR N-PSDB; ABK54192.  
XX Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -  
XX  
XX Claim 10; Page 100; 128pp; English.  
PS The invention describes assays for identifying a (potential) antibiotic  
XX comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.  
XX  
SQ Sequence 254 AA;  
Query Match 20.7%; Score 275.5; DB 23; Length 254;  
Best Local Similarity 28.1%; Pred. No. 9.2e-20;  
Matches 72; Conservative 58; Mismatches 105; Indels 21; Gaps 5;  
QY 9 LIIDIGNTSIAFALFKDNQVNLFIKMTNMLRYDEVYSFFEEENFDF-----NVNKFVI 62  
Db 3 LVIDVGNTNTVLGVYQDETLLVHHWRLATSRQKTEDEYAMTVRSFLDHAGLQFDIDGIVI 62  
QY 63 SSVVPILNETKKNVIFSFFKIKPLFI-----GFDLNYDLTFNPKYKSDKFLGSDVFANL 116  
Db 63 SSVVPMMFSLQMKCKKYFHVTPMIIGPGIKTGLNIKYD--NPKE-----VGADRIVNA 114  
QY 117 VAAIENYSFENVLVLDGTACTIFAVSRQDGILGGIINSGPLINFNLSLDNAYLIKKEPFI 176  
Db 115 VAAIELYGY-P-AIVVDFGTATTTCYLINEKKQYAGGVIAPGIMISTEALYHRASKLPRIE 173  
QY 177 STPNLLERTTSGSVNSGLFYQYKYLIEGVYRDIKQYKKKFNLIITGGNADLILSLIEI 236  
Db 174 AKPKQVVGVTNTIDSMQSGIFYGYVSQVDGVVVKRMKAQAESEPKVIATGGLAKLIGTESE 233  
QY 237 EFIFNIHLTVEGVRIL 252  
Db 234 IDVIDSFLTLKGLQLI 249  
RESULT 5  
AAU01243  
ID AAU01243 standard; Protein; 258 AA.  
XX  
AC AAU01243;  
XX  
DT 18-JUL-2001 (first entry)  
XX  
DE B. subtilis novel pantothenate kinase encoded by the gene coax.  
XX Pantothenate kinase; coax; pantothenate biosynthesis; vitamin B5;  
KW nutritional supplement; panto-compound; pantoate.  
XX  
OS Bacillus subtilis.  
XX  
XX WO200121772-A2.  
XX  
XX 29-MAR-2001.  
XX  
XX 21-SEP-2000; 2000WO-US25993.  
XX PF 21-SEP-1999; 99US-0400494.  
XX PR 07-JUN-2000; 2000US-0210072.  
XX PR 28-JUL-2000; 2000US-0221836.  
XX PR 24-AUG-2000; 2000US-0227860.  
XX

PA (OMNI-) OMNIGENE BIOPRODUCTS.  
XX Yocum RR, Patterson TA, Hermann T, Pero JG;  
PI WPI; 2001-218644/22.  
XX N-PSDB; AAS00984.  
DR New recombinant microorganism which overexpress a Bacillus subtilis  
XX pantothenate biosynthetic enzyme, useful for the high yield production  
PT of panto-compounds such as pantothenate and pantoate -  
PT Example 14; Fig 23; 292pp; English.  
XX The sequence represents a novel B. subtilis pantothenate kinase (encoded  
CC by gene coax), an enzyme of the pantothenate biosynthetic pathway.  
CC Pantothenate, also known as vitamin B5, is used as a nutritional  
CC supplement in mammals and humans. The invention concerns methods of  
CC producing recombinant microorganisms overexpressing at least one Bacillus  
CC subtilis pantothenate biosynthetic enzyme. The microorganisms and methods  
CC of producing them are useful for producing a panto-compound such as  
CC pantothenate or pantoate, which is a nutritional requirement for  
CC livestock and humans. The methods are also useful for the identification  
CC of pantothenate kinase modulators. Panto-compounds are produced at a  
CC significantly higher yield than prior art methods and can be produced  
CC independent of the need to feed precursors which decreases expense.  
XX  
SQ Sequence 258 AA;  
Query Match 20.4%; Score 271.5; DB 22; Length 258;  
Best Local Similarity 27.3%; Pred. No. 2.4e-19;  
Matches 70; Conservative 57; Mismatches 108; Indels 21; Gaps 5;  
QY 9 LIIDIGNTSIAFALFKDNQVNLFIKMKTNLMRLRYDEVYSFFEENFDN-----VNKVF 62  
Db 3 LVIDVGNTNTVLGVYHDGKLEYHWRHRIETSRHKTEDEFGMILRSFLDHSGLMFEQIDGIII 62  
QY 63 SSVVPILNETKKNVIFSFYKIKPLFI-----GFDLNYDLTFNPKSKDKFLGSDVFANL 116  
Db 63 SSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNLIKID---NPKE-----VGADRIVNA 114  
QY 117 VAAIENYSFENVLVVDLGTACTIFAVSRQDGILGGIINSGLPLNFNSLLDNAYLIKKEPPI 176  
Db 115 VAAIHLYG-NPLIVDFGTATTYCYIDENKQYMGGAIAPIGTISTEALYSRAAKLPRIE 173  
QY 177 STPNLLERTTSGVNSGLFYQYKYLIEGVYRDIKQYKKFNLIITGGNADLILSLIEI 236  
Db 174 TRPDNIIGKNTVSAMQSGILFGYVQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDC 233  
QY 237 EFIFNIHLTVEGVRI 252  
Db 234 IDIVDPFLTKGLELI 249  
RESULT 6  
AAU91149  
ID AAU91149 standard; Protein; 258 AA.  
XX AAU91149;  
AC  
XX 05-JUN-2002 (first entry)  
DT  
XX Bacillus subtilis pantothenate kinase Coax.  
DE  
XX Pantothenate kinase; Coax; antilbolic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX  
OS Bacillus subtilis.  
XX WO200216601-A2.  
\*PN  
XX 28-FEB-2002.  
PD  
XX 24-AUG-2001; 2001WO-US26531.  
PF

XX 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
PA Yocum RR, Patterson TA;  
PI WPI; 2002-269358/31.  
XX N-PSDB; ABK54168.  
DR Identifying potential antibiotic or antimicrobial agent, comprises  
XX contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -  
PT Claim 10; Page 67-68; 128pp; English.  
PS The invention describes assays for identifying a (potential) antibiotic  
XX comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.  
XX  
SQ Sequence 258 AA;  
Query Match 20.4%; Score 271.5; DB 23; Length 258;  
Best Local Similarity 27.3%; Pred. No. 2.4e-19;  
Matches 70; Conservative 57; Mismatches 108; Indels 21; Gaps 5;  
QY 9 LIIDIGNTSIAFALFKDNQVNLFIKMKTNLMRLRYDEVYSFFEENFDN-----VNKVF 62  
Db 3 LVIDVGNTNTVLGVYHDGKLEYHWRHRIETSRHKTEDEFGMILRSFLDHSGLMFEQIDGIII 62  
QY 63 SSVVPILNETKKNVIFSFYKIKPLFI-----GFDLNYDLTFNPKSKDKFLGSDVFANL 116  
Db 63 SSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNLIKID---NPKE-----VGADRIVNA 114  
QY 117 VAAIENYSFENVLVVDLGTACTIFAVSRQDGILGGIINSGLPLNFNSLLDNAYLIKKEPPI 176  
Db 115 VAAIHLYG-NPLIVDFGTATTYCYIDENKQYMGGAIAPIGTISTEALYSRAAKLPRIE 173  
QY 177 STPNLLERTTSGVNSGLFYQYKYLIEGVYRDIKQYKKFNLIITGGNADLILSLIEI 236  
Db 174 TRPDNIIGKNTVSAMQSGILFGYVQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDC 233  
QY 237 EFIFNIHLTVEGVRI 252  
Db 234 IDIVDPFLTKGLELI 249  
RESULT 7  
AAU91154  
ID AAU91154 standard; Protein; 255 AA.  
XX AAU91154;  
AC  
XX 05-JUN-2002 (first entry)  
DT  
XX Geobacter sulfurreducens pantothenate kinase Coax.  
DE  
XX Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX  
OS Geobacter sulfurreducens.  
XX WO200216601-A2.  
PN  
XX 28-FEB-2002.  
PD





```

XX 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
XX
XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX Yocum RR, Patterson TA;
PI
XX WPI; 2002-269358/31.
DR N-PSDB; ABK54193.
XX
XX Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein -
XX
XX Claim 10; Page 101-102; 128pp; English.
PS
XX The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX
XX Sequence 258 AA;
SQ
Query Match 19.3%; Score 256.5; DB 23; Length 258;
Best Local Similarity 26.3%; Pred. NO. 8.1e-18;
Matches 67; Conservative 59; Mismatches 108; Indels 21; Gaps 5;
QY 10 IIDIGNTSIAFALEKDNQVNLFIKMKTNLMRLRYDEVYSFFEEFDF-----NVNKFVIS 63
DB 4 VLDVGNTNTVLGVYDDELKHHWRIETSRKTEDEYGMKALLNHVGLQFSDIRGIIS 63
QY 64 SVVPILNETFRKNVIFSFFKIKPLFI-----GFDLNYDLTFNPKSKDFLLGSDVFANLV 117
DB 64 SVVPPIMFALERMCLKFKPLIVPGPIKTLGDIKYD---NPRE-----VGADRIVNAV 115
QY 118 AAIENYSFENVLVVDLGTACTIFAVSRQDGLGGIINSGPLINFNSLLDNAYLIKKEPIS 177
DB 116 AGIHLYG-SPLIIVDFGTATTTCYINERHKOYMGGAIAPIGIMISTEALFARAALPRIEIA 174
QY 178 TPNLLERTSGSVNSGLFYQYKYLIEGVYRDIKQMKKFNLIITGGNADLILSLIEIE 237
DB 175 RPDDIIGKNTVSAMQAGILYGYVQVEGIVSRMKAKSKIPPKVIATGGLAPLIASESDII 234
QY 238 FIFNIHLTVEGVRIIL 252
DB 235 DVVDPFLTLTKLKL 249
RESULT 10
ID ABB47661 standard; Protein; 259 AA.
XX
XX AC ABB47661;
XX
XX 05-FEB-2002 (first entry)
XX Listeria monocytogenes protein #365.
XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW vitamin B12; bacterial infection; disease.
XX
XX Listeria monocytogenes.
XX
XX WO200177335-A2.
XX
XX 18-OCT-2001.
```

```

XX 11-APR-2001; 2001WO-FR01118.
XX
XX 11-APR-2000; 2000FR-0004629.
XX (INSP ) INST PASTEUR.
XX
XX Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI Maduenio E, De Pablo B, Wehland J, Kaerst U, Entian K, Hauf J;
PI Rose M, Voss H;
XX
XX WPI; 2002-010914/01.
DR
XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT and prevention of Listeria and related bacterial infections, and
PT related polypeptides -
XX
XX Claim 6; SEQ ID NO 366; 192pp; French.
XX
XX The present invention relates to the genome sequence of Listeria
CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
CC it are useful for selecting probes and primers for detecting genes in L.
CC monocytogenes and related organisms, and for studying genetic
CC polymorphisms and other genomes. The present sequence is a protein
CC encoded by the genome sequence of the present invention. Proteins
CC expressed from the genome sequence are useful for raising specific
CC antibodies, identification of L. monocytogenes and related organisms, and
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC B12. The genome sequence and proteins encoded by it are also useful for
CC selecting compounds that regulate gene expression and cell replication
CC and modulate L. monocytogenes-related diseases. In addition, the genome
CC sequence and proteins encoded by it are useful in pharmaceutical and
CC vaccine compositions for the treatment or prevention of infections by L.
CC monocytogenes and related organisms.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 259 AA;
SQ
Query Match 18.7%; Score 248.5; DB 23; Length 259;
Best Local Similarity 27.7%; Pred. NO. 5.3e-17;
Matches 70; Conservative 53; Mismatches 115; Indels 15; Gaps 5;
QY 9 LIIDIGNTSIAFALEKDNQVNLFIKMKTNLMRLRYDE----VYSFFE--ENFDENVKVF 62
DB 3 LVIDVGNTNCTGVYEKQKLLKHWRMTDRHRTSDELGMTVLNFFSYANLTPSDIQGI 62
QY 63 SSVVPILNETFRKNVIFSFFKIKPLFIIGFDLNYDLTF---NPYKSDKFLLGSDVFANLVA 119
DB 63 SSVVPPIMHAMETMVCVRYFNIRPLIVPGPIKTLGTLNKLKVDNPRE-----IGSDRIVNAVA 117
QY 120 IENYSFENVLVVDLGTACTIFAVSRQDGLGGIINSGPLINFNSLLDNAYLIKKEPIS 179
DB 118 SEEYGT-TPVIVDFGTATTTCYIDESGVYQGGAIAPGIMISTEALYNRAAKLPRVDIAES 176
QY 180 NNLLERTSGSVNSGLFYQYKYLIEGVYRDIKQMKKFNLIITGGNADLILSLIEIEFI 239
DB 177 SQIIGKSTVSSMQAGIFYGVGQCEGIIAEMKKQSNASPVVATGGLARMIKSSAVDI 236
QY 240 FNIHLTVEGVRIIL 252
DB 237 LDPFLTKGLELL 249
RESULT 11
AAU91163
ID AAU91163 standard; Protein; 233 AA.
```







XX 05-JUN-2002 (first entry)  
DT XX  
DE XX  
DE XX  
DE XX  
KW Pantothenate kinase (Coax) #11.  
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX  
OS Caulobacter crescentus.  
XX  
PN WO200216601-A2.  
XX  
PD 28-FEB-2002.  
XX  
PF 24-AUG-2001; 2001WO-US26531.  
XX  
PR 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX  
A (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX  
PI Yocum RR, Patterson TA;  
XX  
DR WPI; 2002-269358/31.  
DR N-PSDB; ABK54194.  
XX  
PT Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -  
XX  
PS Claim 10; Page 102-103; 128pp; English.  
XX  
CC The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.  
XX  
SQ Sequence 260 AA;  
  
Query Match 14.2%; Score 189.5; DB 23; Length 260;  
Best Local Similarity 26.9%; Pred. No. 5.5e-11;  
Matches 67; Conservative 41; Mismatches 130; Indels 11; Gaps 4;  
  
QY 9 LIIDIGNTSIAFALFKDNQVNLFIKMKTNLMRLRYDEVYSFPEENFDN-----VNKVF 62  
Db 3 LAIEQGNNTMTFAIHGASWVAQWRSATESRTRTADEVYVWLSQLLSMQGLGFRAIDAVII 62  
  
QY 63 SSVVPILNETFKNVIFSEFKIKPLFIGEDLNYDLTFNPKYKDKFLLGSDVFANLVAATEN 122  
Db 63 SSVVPQSIFNRLNLSRRYFNVEPLVIGENAKLGIDVRIEKPSE--AGADRLVNAIGAAMV 120  
  
QY 123 YSFENVLVVDLGTACTIFAVSRQDGIILGGIINSGPLINFSLLDNAYLIKKEPISIP--N 180  
Db 121 YP-GPLVVIDSGTATTFDVAADGAFEGGIIAPGINLSMQALHEAAKLPRIAIQRPAGN 179  
  
QY 181 NLLERTTSGVNSGLFYQYKYLIEGVYRDIKQYKKKNLIIITGGNADLILSLIEIEIF 240  
Db 180 RIVGTDVTSAMQSGVFWGYISLIEGLVARIKAEERGEPTVIATGGVASLFEGATDSIDHF 239  
  
QY 241 NIHLTVEGV 249  
Db 240 DSDLTIRGL 248

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:08:46 ; Search time 11.1246 Seconds  
(without alignments)  
2373.336 Million cell updates/sec

Title: US-09-813-453A-41  
Perfect score: 1229  
Sequence: 1 MSFNLIVDQGNACKVAFVR.....ILHDPDLVLLGLNRILEYNV 244

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues  
Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :		Published_Applications_AA.*	
1:	/cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*		
2:	/cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*		
3:	/cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*		
4:	/cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*		
5:	/cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*		
6:	/cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*		
7:	/cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*		
8:	/cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*		
9:	/cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*		
10:	/cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*		
11:	/cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*		
12:	/cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*		
13:	/cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*		
14:	/cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1229	100.0	244	9	US-09-813-453A-41
2	224	18.2	258	9	US-09-813-453A-6
3	221	18.0	257	9	US-09-813-453A-53
4	210.5	17.1	258	9	US-09-813-453A-49
5	208	16.9	255	9	US-09-813-453A-7
6	207.5	16.9	262	9	US-09-813-453A-45
7	207	16.8	258	9	US-09-813-453A-2
8	205.5	16.7	265	9	US-09-813-453A-4
9	201.5	16.4	256	9	US-09-813-453A-55
10	198	16.1	260	9	US-09-813-453A-51
11	197.5	16.1	254	9	US-09-813-453A-47
12	195.5	15.9	248	9	US-09-813-453A-20
13	184	15.0	219	9	US-09-813-453A-57
14	183.5	14.9	241	9	US-09-813-453A-63
15	181.5	14.8	242	9	US-09-813-453A-65
16	181	14.7	249	9	US-09-813-453A-61
17	178	14.5	249	9	US-09-813-453A-70
18	175.5	14.3	233	9	US-09-813-453A-17
19	174	14.2	212	9	US-09-813-453A-59

20	172.5	14.0	273	9	US-09-813-453A-10	Sequence 10, Appl
21	167.5	13.6	250	9	US-09-813-453A-3	Sequence 3, Appl
22	166	13.5	460	9	US-09-813-453A-39	Sequence 39, Appl
23	163	13.3	267	9	US-09-813-453A-15	Sequence 15, Appl
24	162	13.2	246	9	US-09-813-453A-9	Sequence 9, Appl
25	161	13.1	592	9	US-09-813-453A-43	Sequence 43, Appl
26	156	12.7	592	9	US-09-813-453A-22	Sequence 22, Appl
27	153	12.4	272	9	US-09-712-363-276	Sequence 276, App
28	153	12.4	272	9	US-09-813-453A-5	Sequence 5, Appl
29	152	12.4	257	9	US-09-813-453A-13	Sequence 13, Appl
30	135	11.0	262	9	US-09-813-453A-11	Sequence 11, Appl
31	132.5	10.8	229	9	US-09-813-453A-12	Sequence 12, Appl
32	131	10.7	262	9	US-09-813-453A-8	Sequence 8, Appl
33	129.5	10.5	223	9	US-09-895-913A-74	Sequence 74, Appl
34	129.5	10.5	223	9	US-09-813-453A-14	Sequence 14, Appl
35	129.5	10.5	223	9	US-09-813-453A-67	Sequence 67, Appl
36	126.5	10.3	209	9	US-09-813-453A-21	Sequence 21, Appl
37	95.5	7.8	468	10	US-09-815-242-5530	Sequence 5530, Ap
38	95.5	7.8	468	10	US-09-815-242-12355	Sequence 12355, A
39	95.5	7.8	468	10	US-09-815-242-12934	Sequence 12934, A
40	91.5	7.4	783	9	US-09-738-626-5017	Sequence 5017, Ap
41	86	7.0	650	9	US-09-993-241-2	Sequence 2, Appl
42	86	7.0	650	10	US-09-993-038-2	Sequence 2, Appl
43	86	7.0	653	9	US-10-206-443-2	Sequence 2, Appl
44	86	7.0	1140	9	US-09-974-973-19	Sequence 19, Appl
45	86	7.0	1140	9	US-09-738-626-4265	Sequence 4265, Ap

ALIGNMENTS

RESULT 1  
US-09-813-453A-41  
; Sequence 41, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn ver. 2.0  
; SEQ ID NO 41  
; LENGTH: 244  
; TYPE: PRT  
; ORGANISM: Porphyromonas gingivalis  
US-09-813-453A-41

Query Match		100.0%	Score 1229;	DB 9;	Length 244;
Best Local Similarity		100.0%	Pred. No. 3.3e-108;		
Matches 244;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MSFNLIVDQGNACKVAFVRNNSIESISFLPGKAGQALSHLVAPHRFDKAIYSSVGLPDE	60		
Db	1	MSFNLIVDQGNACKVAFVRNNSIESISFLPGKAGQALSHLVAPHRFDKAIYSSVGLPDE	60		
QY	61	EAEAIVRSCAAASLMMGTETPVLRLOYDRRTLGADRLAAVVGAAHSLYPNTELLVIDAGT	120		
Db	61	EAEAIVRSCAAASLMMGTETPVLRLOYDRRTLGADRLAAVVGAAHSLYPNTELLVIDAGT	120		
QY	121	AITYERVSAEGIYLGGNISPGHLRKFALHFTGRPLIDPSGISPKIAEYGSSTEEAIT	180		
Db	121	AITYERVSAEGIYLGGNISPGHLRKFALHFTGRPLIDPSGISPKIAEYGSSTEEAIT	180		
QY	181	AGVIHGLAGEIDRYIDDLHAKEGRSVILTTGGDANYLARIIRSGILIHDPDLVLLGLNRIL	240		





Db 216 KVIATGGLAPLIAS--ESDIIDVDPFLTLTGLKLLYEKN 253

RESULT 5  
US-09-813-453A-7  
; Sequence 7, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 255  
; TYPE: PRT  
; ORGANISM: Geobacter sulfurreducens  
US-09-813-453A-7

Query Match 16.9%; Score 208; DB 9; Length 255;  
Best Local Similarity 34.1%; Pred. No. 8.2e-12;  
Matches 59; Conservative 34; Mismatches 64; Indels 16; Gaps 7;  
QY 77 GTETPVPLRLQYDR-RTLGADRLAAVVGAAHSLYPNTELLVIDAGTAITYERVSAEGIYLG 135  
Db 91 GIKTGMP--IQYDNPREGVADRIVNAVAGYEKY-RTSLIIVDFGTATFDYVNRKGEYCG 147  
QY 136 GNISPLHLRFKALHFTGRLPLID---PSGISPKIAEYGSSTEEAITAGVIHGLAGEID 192  
Db 148 GAIAPGLVISTEALFQRASKLPRVDIIRPSAILAR-----NTVNSMQAGIYYGYVGLVD 201  
QY 193 RYIDDLHAKGRSA--VILTGGDANYLARIIRSGILIHDPDLVLLGLNRILEYN 243  
Db 202 EIVTRMKA-ESKDAPRVIAATGGLASLIAPESKTIEAVEEYLTLEGLRILYERN 253

RESULT 6  
US-09-813-453A-45  
; Sequence 45, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 45  
; LENGTH: 262  
; TYPE: PRT  
; ORGANISM: Bacillus anthracis  
US-09-813-453A-45

Query Match 16.9%; Score 207.5; DB 9; Length 262;  
Best Local Similarity 30.8%; Pred. No. 9.5e-12;  
Matches 72; Conservative 44; Mismatches 73; Indels 45; Gaps 14;

QY 25 ESISFLPGKAGQALSHLVAPHRFDKAIYSSVGLPDEEAEAIVRSC-----AAASLMMGTET 80  
Db 50 EGLSFEDVK-GIIVSSVVPPIMF-----ALERMCEKYFKIKPLVVVPGGI 92  
QY 81 PVPLRLQYDR-RTLGADRLA-AVVGAAHSLYPNTELLVIDAGTAITYERVSAEGIYLGNI 138  
Db 93 KTGLNIKYENPREVGADRIVNAVAGIH-LY-GSPLIIVDFGTATTTCYINEEKHYMGVI 150  
QY 139 SPGLHLRFKALHFTGRLPLID---PSGISPKIAEYGSSTEEAITAGVIHGLAGEIDRYI 195  
Db 151 TPGIMISAEALYSRAAKLPRIETKPSV-----VGKNTVSAMQSGILYGYVGQVEGIV 204  
QY 196 DDL--HAKEGRSAVILTGGDANYLARI--RSGIL--IHPDLVLLGLNRILEYN 243  
Db 205 KRMKEAKQ-EPKVIATGG---LAKLISEESNVIDVDPFLTLKGLYMLYERN 253

RESULT 7  
US-09-813-453A-2  
; Sequence 2, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 258  
; TYPE: PRT  
; ORGANISM: Bacillus subtilis  
US-09-813-453A-2

Query Match 16.8%; Score 207; DB 9; Length 258;  
Best Local Similarity 26.4%; Pred. No. 1e-11;  
Matches 71; Conservative 49; Mismatches 101; Indels 48; Gaps 10;

QY 5 LIVDQGNSACKVAFVRNNSIESISFLPGKAGQALSHLVAPHRFD-----KAIYSSVGL 57  
Db 3 LVIDVGNT-----NTVLGV-YHDGKLEYHWRITSRHKTEDEFGMILRSLFDHSG 52  
QY 58 PDEEAEAIVRSCAAASLM-----MGTEPVPPLRLQYDR-RTLGADRLA 99  
Db 53 MFEQIDGIIISVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIV 112  
QY 100 AVVGAHSLYPNTELLVIDAGTAITYERVSAEGIYLGNIISPLHLRFKALHFTGRLPLI 159  
Db 113 NAVAAIHLYGN-PLIVVDFGTATTTCYIDENKQYMGGAITAPGITISTEALYSRAAKLPRI 171  
QY 160 D---PSGISPKIAEYGSSTEEAITAGVIHGLAGEIDRYIDDL--HAKEGRSAVILTGGDA 214  
Db 172 EITRPDNI-----IGKNTVSAMQSGILFYGVGVQVEGIVKRMKQAKQDLK-VIATGG 224  
QY 215 NYLARIIRSGILIHDPDLVLLGLNRILEYN 243  
Db 225 PLIANESDCIDIVDPFLTLKGLELIYERN 253

RESULT 8  
US-09-813-453A-4  
; Sequence 4, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.

```

; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 265
; TYPE: PRT
; ORGANISM: Streptomyces coelicolor
US-09-813-453A-4

Query Match      16.7%; Score 205.5; DB 9; Length 265;
Best Local Similarity 33.5%; Pred. No. 1.5e-11;
Matches 59; Conservative 36; Mismatches 64; Indels 17; Gaps 7;

QY 77 GTEPTVPLRLQYDRRTLGADRLAAVVGAAHSLYPNTELLVIDAGTAITYERSVAEGYILGG 136
| : | | : : : : | | | | : : | | : : | | | | : : | | | | : : | | | |
Db 96 GVKTCVPILTDHPKE-VGADRINAAVAVELY-GGPAIVVDFTGATTFDAVSARGEYIGG 153
| : | | : : | | : : | | : : | | : : | | : : | | : : | | : : | | : :
QY 137 NISPGHLRFKALHLFTGRLPLIDPDSGISPCKIAEYGSSTEEAITAGVIHGLAGEID---- 192
| : | | : : | | : : | | : : | | : : | | : : | | : : | | : : | | : :
Db 154 VIAPGIEISVEALGVKGAQLRKIE---VARPRSVIGKNTVEAMQSGIVYGFAGQVDGVVN 210
| : | | : : | | : : | | : : | | : : | | : : | | : : | | : : | | : :
QY 193 RYIDDLHAKEGRSAVILTGADANYLARI--RSGIL--IHPDLVLLGLNRILEYV 244
| : | | : : | | : : | | : : | | : : | | : : | | : : | | : : | | : :
Db 211 RMARELADDPDDVTVIATGG----LAPMVLGESSVIDEHEPWLTLMLGLRLVYERNV 262
| : | | : : | | : : | | : : | | : : | | : : | | : : | | : : | | : :

RESULT 9
US-09-813-453A-55
; Sequence 55, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 55
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-09-813-453A-55

Query Match      16.4%; Score 201.5; DB 9; Length 256;
Best Local Similarity 27.0%; Pred. No. 3.4e-11;
Matches 75; Conservative 44; Mismatches 94; Indels 65; Gaps 11;

QY 5 LIVDQGNACKVAFVRNNSI-----ESISFLPKAGQALSHLVAPHREDKATYSS 54
| : | | : : : : | | : : | | : : | | : : | | : : | | : : | | : :
Db 3 LVFDVGNNTNMVLGIYKGDVLVNYWRIKTDREKTS---DEYGILISNL-----FD---YDN 51
| : | | : : | | : : | | : : | | : : | | : : | | : : | | : : | | : :
QY 55 VGLPDEEAEAI-----VRSCAAASLMMGTETPVPLRLQYDR-RTLGDRL 98
| : | | : : | | : : | | : : | | : : | | : : | | : : | | : : | | : :
Db 52 VNISDIDDVVISSVVPVNMHSLNFCIKYCKQKPLIVPGIKTGLNLIKYNPKQVGADRI 111
| : | | : : | | : : | | : : | | : : | | : : | | : : | | : : | | : :
QY 99 AAVVGAHSLYPNTELLVIDAGTAITYERSVAEGYILGNNISPGHLRFLKALHLFTGRPL 158
| : | | : : | | : : | | : : | | : : | | : : | | : : | | : : | | : :
Db 112 VNAVAGIEKYGAPSLV-DFGTATTFCAISEKGEYLGGTIAPGIKISSEALFQSASKLPR 170
| : | | : : | | : : | | : : | | : : | | : : | | : : | | : : | | : :

```

```

QY 159 IDPSGISPKIAEYG---SSTEEAITAGVIHGLAGEIDRYI-----DDLHAKEGRS 205
| : | | : : | | : : | | : : | | : : | | : : | | : : | | : :
Db 171 VE-----LAKPGMTICKSTVSAMQSGIIYGVYGLVDKIIISIMKKELNCDDV----- 216
| : | | : : | | : : | | : : | | : : | | : : | | : : | | : :
QY 206 AVILTGADANYLARIIRSGILIHDPDLVLLGLNRILEYN 243
| | | | | | : | : | | : | | | | | | | | | |
Db 217 KVIATGGLAKLIASETKSIDYVDGFLTLEGRLRIYEKN 254
| | | | | | : | : | | : | | | | | | | | | |

RESULT 10
US-09-813-453A-51
; Sequence 51, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Caulobacter crescentus
US-09-813-453A-51

Query Match      16.1%; Score 198; DB 9; Length 260;
Best Local Similarity 33.7%; Pred. No. 7.4e-11;
Matches 59; Conservative 34; Mismatches 72; Indels 10; Gaps 6;

QY 74 LMMGTETPVPLRLQYDRRT-LGADRLAAVVGAAHSLYPNTELLVIDAGTAITYERSVAEGI 132
| : | | : : : : : : | | | | : : | | : : | | : : | | : : | | : :
Db 86 LVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAMVYPG-PLVVIDSGTATTFDIVAADGA 144
| : | | : : | | : : | | : : | | : : | | : : | | : : | | : :
QY 133 YLGGNISPGHLRFLKALHLFTGRPLI---DPSGISPKIAEYGSSTEEAITAGVIHGLAG 189
| : | | : : | | : : | | : : | | : : | | : : | | : : | | : :
Db 145 FEGGIIAPGINLSMQALHEAAAKLPRIAIQRPAG--NRIV--GTDTVSAMQSGVFWGYIS 200
| : | | : : | | : : | | : : | | : : | | : : | | : : | | : :
QY 190 EIDRYIDDLHAKEGR-SAVILTGADANYLARIIRSGILIHDPDLVLLGLNRILEYN 243
| : | | : : | | : : | | : : | | : : | | : : | | : : | | : :
Db 201 LIEGLVARIKAERGEPTMTVIATGGVASLFEAGATDSIDHFDSDLTIRGLLEIYRRN 255
| : | | : : | | : : | | : : | | : : | | : : | | : : | | : :

RESULT 11
US-09-813-453A-47
; Sequence 47, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Bacillus halodurans

```









QY 44 PHRFDKAIYSSVGLPDEAEAIVR-----SCAAASLMMGTETPVPLRLQYDRR 91  
Db 73 PPDVWISVWAGPALRDALCARIARVFDGARLRIVASEAAAGLRNGYRDP----- 123  
QY 92 TLGADRLAAVVGHAHSLYPNTTELLVIDAGTAITYERSAEGIYLGNIISPLGLHLPKALHL 151  
Db 124 QLGTDRWVGAVGARHAWPDTALLVLTAGTATTLDIVAPDGRFAGGLLPLGLTLMRALSR 183  
QY 152 FTGRLPLIDPSGISPK-----IAEYGSSTEEAITAGVIHGLAGEIDRYIDDLHAK-E 202  
Db 184 NTAQLPEIDIGYLAARDDAQAPADVPWSADNTQDAIALGCVTAQAAGIAQTWQALQAQYP 243  
QY 203 GRSVILTGGDANYLARIIRSGLIHHPDLYLLGL 236  
Db 244 GPYRCVLSGGARAALAPHLRMPFQMHDNLVLLGL 277

RESULT 2

Q9F985 PRELIMINARY; PRT; 258 AA.  
AC Q9F985;  
DT 01-MAR-2001 (TReMBLrel. 16, Created)  
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Putative 32 kDa replication protein.  
OS Bacillus stearothermophilus.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Geobacillus.  
OX NCBI\_TaxID=1422;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-V;  
RA Vasquez C., Pichuanes S., Saavedra C.;  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF198621; AAG28531.1;  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMs; TIGR00671; baf; 1.  
SQ SEQUENCE 258 AA; 28101 MW; 507B55D695095855 CRC64;

Query Match 18.1%; Score 222.5; DB 2; Length 258;  
Best Local Similarity 31.1%; Pred. No. 2.4e-10;  
Matches 68; Conservative 39; Mismatches 81; Indels 31; Gaps 9;

QY 49 KAIYSSVGLPDEAEAIVRSCAAASLMMGTET-----TPV-----PLRLQYDR- 90  
Db 44 KALLNHVGLQFSDIDGIIISVVPPIMFALERNCLKFHFKPIIVGPGIKTGLNIKYDNP 103  
QY 91 RTLGAADRLA-AVVGHAHSLYPNTTELLVIDAGTAITYERSAEGIYLGNIISPLGLHLPKAL 149  
Db 104 REVGADRIVNAVAGIH-LY-GSPLIIVDFGTATYCYINEHKQYMGGAIAPIGIMISTEAL 161  
QY 150 HLFTGRLPID---PSGISPKIAEYGSSTEEAITAGVIHGLAGEIDRYIDDLHAKEG-RS 205  
Db 162 FARAACKLPRIETARPDDI-----IGKNTVSAMQAGILYGVGVQEGIVSRMKAKSPVPP 215  
QY 206 AVILTGGDANYLARIIRSGLIHHPDLYLLGLNRILEYNV 244  
Db 216 KVIATGGGLASLIASENSVIDIVDPFLTLTGLKILYKENV 254

RESULT 3

Q8YAC5 PRELIMINARY; PRT; 259 AA.  
AC Q8YAC5;  
DT 01-MAR-2002 (TReMBLrel. 20, Created)  
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Hypothetical protein lmo0221.  
GN LMO0221.  
OS Listeria monocytogenes.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;

OC Listeriaceae; Listeria.  
OX NCBI\_TaxID=1639;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-EGD-E / SEROVAR 1/2A;  
RX MEDLINE-21537279; PubMed-11679669;  
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A., Chakraborty T., Baquero F., Berche P., Bloeker H., Brandt P., Daruvar A., Dehoux P., Charbit A., Chetouani F., Couve E., de Daruvar A., Durant L., Dussurget O., Domann E., Dominguez-Bernal G., Duchaud E., Hain T., Hauf J., Jackson D., Gantier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Kurapkat G., Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Medjari H., Madueno E., Maitournam A., Mata Vicente J., Ng E., Purcell R., Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Tierrez A., Remmel B., Rose M., Schlueter T., Simoes N., Wehland J., Cossart P.;  
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;  
RT "Comparative genomics of Listeria species."  
RL Science 294:849-852(2001).  
DR EMBL; AL591974; CAD00748.1;  
DR MEROPS; M41.009;  
DR ListList; LMO00221;  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMs; TIGR00671; baf; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 259 AA; 28187 MW; F106B049D80025B7 CRC64;

Query Match 17.6%; Score 216; DB 16; Length 259;  
Best Local Similarity 25.4%; Pred. No. 8e-10;  
Matches 69; Conservative 43; Mismatches 106; Indels 54; Gaps 9;

QY 5 LIVDQGSACKVAFVRNNSIESISFLPGKAGQALSH---LVAPHRFDKAI-----YSS 54  
Db 3 LVIDVGNTNCTVGVEYKQL-----LKHWRMTDRHRTSDELGMTVLNFFSY 49  
QY 55 VGLPDEAEAIVRSCAAASLMMGTET-----PVPLRLQYDR-RTLGD 96  
Db 50 ANLTPSDIQIIISVVPPIMHAMETMCVRYFNIRPLIVGPGIKTGLNLKVDNPREIGSD 109  
QY 97 RLAAVVGHAHSLYPNTTELLVIDAGTAITYERSAEGIYLGNIISPLGLHLPKALHLFTGRL 156  
Db 110 RIVNAVAASEEY-GTPVIIVDFGTATTTCYIDESGVYQGAIAPIGIMISTEALYNRAAKL 168  
QY 157 PLIDPSGISPKIAE---YGSSTEEAITAGVIHGLAGEIDRYIDDLHAKEGRSVIL-TG 211  
Db 169 PRVD-----IAESSQIIGKSTVSSMQAGIFYGVGQCEGIIAEMKKQSNASPVVATG 221  
QY 212 GDANYLARIIRSGLIHHPDLYLLGLNRILEYN 243  
Db 222 GLARMIKSSAVDILDPLTLKGLLELYRRN 253

RESULT 4

Q8R7M2 PRELIMINARY; PRT; 255 AA.  
AC Q8R7M2;  
DT 01-JUN-2002 (TReMBLrel. 21, Created)  
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Putative transcriptional regulator, homologs of Bvg accessory factor.  
DE TTE2381.  
GN TTE2381.  
OS Thermoanaerobacter tengcongensis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.  
OX NCBI\_TaxID=119072;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MB4T / JCM11007;  
RX MEDLINE-21992816; PubMed-11997336;  
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J., Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,



```
RESULT 7
Q9X8N6 PRELIMINARY; PRT; 265 AA.
ID Q9X8N6
AC Q9X8N6;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical protein SCO3380.
GN SCO3380 OR SCE94.31C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL049628; CAB40880.1; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00671; baf; 1.
W Hypothetical protein.
SQ SEQUENCE 265 AA; 28203 MW; E34FA2ABBA946969 CRC64;

Query Match 16.7%; Score 205.5; DB 16; Length 265;
Best Local Similarity 33.5%; Pred. No. 6e-09;
Matches 59; Conservative 36; Mismatches 64; Indels 17; Gaps 7;

QY 77 GTETPVPLRLQYDRRTLGAADRLAAVVGASHLYPNTLLVIDAGTAITYERVSAEGIYLG 136
Db 96 GVKTGVPILTDHPKE-VGADRIINAVAAVELY-GGPAIVDFGTATFDVARSAGEYIGG 153
QY 137 NISGLHLRFKALHLFTGRPLPLIDPSGISPKIAEYGSSTEEAITAGVIHGLAGEID- 192
Db 154 VIAPGIEISVEALGVKGAQLRKIE---VARPRSVIGKNTVEAMQSGIVYGFAGQVDGVN 210
QY 193 RYIDDLHAKGRSAVILTTGGDANYLARI--RSGIL--IHPDLVLLGLNRLILEYNV 244
Db 211 RMARELADDDVDTVIATGG----LAPMVLGESSVIDEHEPWLTLMLGLRLVYERNV 262

RESULT 8
Q9A6Z1 PRELIMINARY; PRT; 261 AA.
ID Q9A6Z1
AC Q9A6Z1;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Transcriptional activator, putative, Baf family.
GN CC1935.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CBI5;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005867; AAK23910.1; -.
DR TIGR; CC1935; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00671; baf; 1.
KW Complete proteome.
SQ SEQUENCE 261 AA; 27965 MW; C19E60D7B0714EF5 CRC64;

Query Match 16.1%; Score 198; DB 16; Length 261;
Best Local Similarity 33.7%; Pred. No. 2.4e-08;
Matches 59; Conservative 34; Mismatches 72; Indels 10; Gaps 6;

QY 74 LMMGTETPVPLRLQYDRRT-LGADRLAAVVGASHLYPNTLLVIDAGTAITYERVSAEGI 132
Db 87 LVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAMVYPG-PLVVIDSGTATTFDIVAADGA 145
QY 133 YLGNISPGHLRFKALHLFTGRPLI---DPSGISPKIAEYGSSTEEAITAGVIHGLAG 189
Db 146 FEGGIIAPGINLSMQALHEAAKLPRIAIQRPAG--NRIV--GTDVTSAMQSGVFWGYIS 201
QY 190 EIDRYIDDLHAKGR-SAVILTTGGDANYLARIIRSGILTHPDLVLLGLNRLILEYN 243
Db 202 LIEGLVARIKAERGEPMTVIATGGVASLFEGATDSIDHFDSDLTIRGLLEIYRRN 256

RESULT 9
Q9KGH5 PRELIMINARY; PRT; 254 AA.
ID Q9KGH5
AC Q9KGH5;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical protein BH0086.
GN BH0086.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hiramata C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001507; BAB03805.1; -.

```



DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMS; TIGR00671; baf; 1.  
KW. Hypothetical protein; Complete proteome.  
SQ SEQUENCE 254 AA; 27907 MW; 0E3B557BA7DAC176 CRC64;  
  
Query Match 16.1%; Score 197.5; DB 16; Length 254;  
Best Local Similarity 27.9%; Pred. No. 2.5e-08;  
Matches 60; Conservative 44; Mismatches 88; Indels 23; Gaps 6;  
  
QY 49 KAIYSSVGLPDEEAIAIVRSCAAASLMMGTE-----TPV-----PLRLQYDR- 90  
Db 44 RSLFDHAGLQFDIDGIVISSVPPMMFSLQCMCKKYFHVTPMIGPIKGTGLNIKYDNP 103  
  
QY 91 RTLGADRLAAVVGHAHSLYPNTTELLVIDAGTAITYERSVSAEGIYLGGNISPGHLRFRKALH 150  
Db 104 KEVGADRIVNAVAIELY-GYPAIVVDFGTATTYCLINEKKQYAGGVIAPIGIMISTEALY 162  
  
Y 151 LFTGRLPLIDPSGISPKIAEYGSSTEEAITAGVIHGLAGEIDRYIDDLHAK-EGRSAVIL 209  
Db 163 HRASKLPRIE---IAKPKQVVGTTNTIDSMQSGIFYGYVSQVDGVVVKRMKAQAESEPKVIA 219  
  
QY 210 TGGDANYLARIIRSGILIHDPDLVLLGLNRILEYNNV 244  
Db 220 TGGLAKLIGTESETIDVIDSFLTKGLQLIYKKNV 254  
  
RESULT 10  
Q9HWC1  
ID Q9HWC1 PRELIMINARY; PRT; 248 AA.  
AC Q9HWC1;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical protein PA4279.  
GN PA4279.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 15692 / PA01;  
RX MEDLINE=20437337; PubMed=10984043;  
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Salier M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
RT opportunistic pathogen."  
RL Nature 406:959-964(2000).  
DR EMBL; AE004843; AAG07667.1; -;  
DR InterPro; IPR004619; Baf.  
DR InterPro; IPR001230; Prenyl\_site.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR PROSITE; PS00294; PRENYLATION; UNKNOWN\_1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 248 AA; 26758 MW; 609D37480899DA3B CRC64;

Query Match 15.9%; Score 195.5; DB 16; Length 248;  
Best Local Similarity 30.4%; Pred. No. 3.6e-08;  
Matches 80; Conservative 33; Mismatches 95; Indels 55; Gaps 12;  
  
QY 5 LIVDQGNACKVAFVRNNSIESISFLPGKAGQALSHLVAPHRFDKAIYSSVGLPDEEA 64  
Db 3 LELDCGNSLIKRWVI-EGAARVA---GGLAESDDALV-----EQLTSQALP----- 46  
  
QY 65 IVRSCAAASLMMGTT-----PVP-----LRLQY-DRRTLGADRLAA 100  
Db 47 -VRACRLVSRSEQETSQVLRLEQLFPVSVLVASSGKQLAGVRNGYLDYQRLGLDRWLA 105

QY 101 VVGAHSLYPNTTELLVIDAGTAITYERSVSAEGIYLGGNISPGHLRFRKALHLFTGRPLID 160  
Db 106 LVAAHHL-AKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQRLTRTHRRIRYDD 164  
  
QY 161 P-----SGISPKIAEYGSSTEEAITAGVIHGLAGEI-DRYIDDLHAKEGRSVILTGGD 213  
Db 165 AEARRALASLOP-----GQATAEAVERGCLLMRLRGFVREQYAMACELLGPDCEIFLTGGD 219  
  
QY 214 ANYLARIIRSGILIHDPDLVLLGL 236  
Db 220 AE-LVRDELAGARIMPDLVFVGL 241  
  
RESULT 11  
Q8XHL5  
ID Q8XHL5 PRELIMINARY; PRT; 259 AA.  
AC Q8XHL5;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical protein CPE2468.  
GN CPE2468.  
OS Clostridium perfringens.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
OC Clostridiales; Clostridiaceae; Clostridium.  
OX NCBI\_TaxID=1502;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=13 / TYPE A;  
RX PubMed=11792842;  
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,  
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;  
RT "Complete genome sequence of Clostridium perfringens, an anaerobic  
RT flesh-eater."  
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).  
DR EMBL; AP003194; BAB82174.1; -;  
DR InterPro; IPR004619; Baf.  
DR InterPro; IPR000515; BPD\_transp.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMS; TIGR00671; baf; 1.  
DR PROSITE; PS00402; BPD\_TRANS\_INN\_MEMBR; UNKNOWN\_1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 259 AA; 28819 MW; 0D5FEA3B7A145E10 CRC64;

Query Match 15.3%; Score 188; DB 16; Length 259;  
Best Local Similarity 29.8%; Pred. No. 1.6e-07;  
Matches 65; Conservative 42; Mismatches 79; Indels 32; Gaps 12;  
  
QY 44 PHRFDKAIYSSVGLPD--EEAEAIVRSC-AAASLMMGTTETPVPLRLQYDR-RTLGADRLA 99  
Db 54 PEDVEGIIISV-VPNIMHSLNVMVRKCFCKEPIVVGPIKGINIKYDNPKEVGADRIV 112  
  
QY 100 AVVGAHSLYPNTTELLVIDAGTAITYERSVSAEGIYLGGNISPGHLRFRKALHLFTGRPLI 159  
Db 113 NAVAAFEKH-KKPMIIDFGTATTTCATTEKGYLGGNICPGIOISADALFERAAKLPR 171  
  
QY 160 D---PSGISPKIAEYGSSTEEAITAGVIHGLAGEIDRYID-----DLHAKEGRSVIL 209  
Db 172 ELEKPKSVICK-----NTVTSMQAGIIYGYIKRVEYIVKRMKKEMMDLGEKE--PFVLA 223  
  
QY 210 TGGDANYLARIIRS--GILHPD--LVLLGLNRILEYN 243  
Db 224 TGG----LAKLVSETDVIDEVRKLTLEGKILYEKN 257

RESULT 12  
Q8YQD7  
ID Q8YQD7 PRELIMINARY; PRT; 276 AA.  
AC Q8YQD7;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical protein Alr3896.

GN ALR3896.  
 OS Anabaena sp. (strain PCC 7120).  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
 OX NCBI\_TaxID=103690;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21595285; PubMed=11759840;  
 RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
 RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,  
 RA Yasuda M., Tabata S.;  
 RT "Complete genomic sequence of the filamentous nitrogen-fixing  
 cyanobacterium Anabaena sp. strain PCC 7120.";  
 RL DNA Res. 8:205-213(2001).  
 DR EMBL; AP003594; BAB75595.1; -.  
 DR InterPro; IPR004619; Baf.  
 DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 276 AA; 30272 MW; 2AD04CA693E56F25 CRC64;

Query Match 14.9%; Score 183; DB 16; Length 276;  
 Best Local Similarity 31.3%; Pred. NO. 4.4e-07;  
 Matches 52; Conservative 33; Mismatches 71; Indels 10; Gaps 4;

QY 82 VPLRLQYDRRTLGADRLAAVVGHAHSLYPNTTELLVIDAGTAITYERSVSAEGYILGNNISPG 141  
 :||| | ||| ||| ||| :||| : ||||| ||| :||| |||  
 Db 110 IPLNNIYP--TLGIDRALALWGAGMSW-GFPVLVIDAGTALTFTAADGGKNLVGGAILPG 166  
 :||| | ||| ||| ||| :||| : ||||| ||| :||| |||  
 QY 142 LHLRFKALHLFTGRPLIDPSGISPKIAEYGSSTEEAITAGVIHGLAGEIDRYIDDLHAK 201  
 :||| | ||| ||| ||| :||| : ||||| ||| :||| |||  
 Db 167 VGLQFASLGGQGTGQLPQVEMEAIKSLPPRFALNTTEAIQSGVIYTLIAGMRDFTTEWLSL 226  
 :||| | ||| ||| ||| :||| : ||||| ||| :||| |||  
 QY 202 EGRSAVILTGDD----ANYLARI---IRSGILHPDLVLLGLNRL 240  
 :||| | ||| ||| ||| :||| : ||||| ||| :||| |||  
 Db 227 FPDGKVAIKGDRILLNLYLQALYPDLAARLIVEPNLIFWGMQTIV 272  
 :||| | ||| ||| ||| :||| : ||||| ||| :||| |||

RESULT 13  
 Q9PC14  
 ID Q9PC14 PRELIMINARY; PRT; 242 AA.  
 AC Q9PC14;  
 DT 01-OCT-2000 (TremBLrel. 15, Created)  
 DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)  
 DE Hypothetical protein xfl1795.  
 GN xfl1795.  
 OS Xylella fastidiosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 Xylella.  
 NCBI\_TaxID=2371;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=9A5C;  
 RX MEDLINE=20365717; PubMed=10910347;  
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
 RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,  
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,  
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,  
 RA Coutauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,  
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,  
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,  
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,

RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,  
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,  
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;  
 RT "The genome sequence of the plant pathogen Xylella fastidiosa.";  
 RL Nature 406:151-159(2000).  
 DR EMBL; AE004001; AAF84603.1; -.  
 DR InterPro; IPR004619; Baf.  
 DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 242 AA; 25658 MW; EDC5DE7C0449C1E9 CRC64;

Query Match 14.8%; Score 181.5; DB 16; Length 242;  
 Best Local Similarity 28.0%; Pred. No. 4.8e-07;  
 Matches 72; Conservative 39; Mismatches 91; Indels 55; Gaps 12;

QY 6 IVDQGSACKVAFVRNNSIESISFLP-----GKAGQALSHLVAPHRFDK 49  
 :||| | ||| ||| ||| :||| : ||||| ||| :||| |||  
 Db 5 LFDLGNRSRKCASLREGVIGPVTVLPYLTETMDAFALQELPRGVAY-LASVAAP----- 58  
 :||| | ||| ||| ||| :||| : ||||| ||| :||| |||  
 QY 50 AIYS-----SVGLPDEEAELVRSCTAAASLMGMGTETPVPLRLOYDR-RTLGAADRLAAVV 102  
 ||| :||| :||| ||| ||| :||| : ||||| ||| :||| |||  
 Db 59 AITHTVLEVLKIHFEQVQVAATVAACAG-----VRIAYAHPERFGVDRLALL 106  
 :||| | ||| ||| ||| :||| : ||||| ||| :||| |||  
 QY 103 GAHSLYPNTTELLVIDAGTAITYERSVSAEGYILGNNISPGHLRFKALHLFTGRPLIDPS 162  
 :||| | ||| ||| ||| :||| : ||||| ||| :||| |||  
 Db 107 GS--YEGNVLVGVGTALTIDLLAANGCHLGRISASPTLMRQALHARAELPL---S 160  
 :||| | ||| ||| ||| :||| : ||||| ||| :||| |||  
 QY 163 GISPKIAEYGSSTEEAITAGVIHGLAGEIDRYIDDLHAKGRSAVILT--GGDANYLARI 220  
 :||| | ||| ||| ||| :||| : ||||| ||| :||| |||  
 Db 161 G--GNYLEFAEDTEDALVSGCNGCAVALIERSLYEAHQRLDQSVRLLLHGGGVASLLPWL 218  
 :||| | ||| ||| ||| :||| : ||||| ||| :||| |||  
 QY 221 IRSGILIH-PDLVLLGL 236  
 :||| | ||| ||| ||| :||| : ||||| ||| :||| |||  
 Db 219 ---GDVVHRPTLVLDGL 232  
 :||| | ||| ||| ||| :||| : ||||| ||| :||| |||

RESULT 14  
 Q8RFE4  
 ID Q8RFE4 PRELIMINARY; PRT; 256 AA.  
 AC Q8RFE4;  
 DT 01-JUN-2002 (TremBLrel. 21, Created)  
 DT 01-JUN-2002 (TremBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)  
 DE Bvg accessory factor.  
 GN FN0761.  
 OS Fusobacterium nucleatum (subsp. nucleatum).  
 OC Bacteria; Fusobacteria; Fusobacterium.  
 OX NCBI\_TaxID=76856;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 25586;  
 RX MEDLINE=21886394; PubMed=11889109;  
 RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,  
 RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,  
 RA Vasleva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,  
 RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,  
 RA Fonstein M., Kyrpides N., Overbeek R.;  
 RT "Genome sequence and analysis of the oral bacterium Fusobacterium  
 nucleatum strain ATCC 25586.";  
 RL J. Bacteriol. 184:2005-2018(2002).  
 DR EMBL; AE010586; AAL94957.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 256 AA; 28601 MW; CDF681127F08914B CRC64;

Query Match 14.5%; Score 178.5; DB 16; Length 256;  
 Best Local Similarity 31.8%; Pred. No. 9.2e-07;  
 Matches 49; Conservative 27; Mismatches 67; Indels 11; Gaps 4;



**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:56 ; Search time 7.128 seconds  
(without alignments)  
1007.182 Million cell updates/sec

Title: US-09-813-453A-41  
Perfect score: 1229  
Sequence: 1 MSFNLIYDQGNACKVAFVR.....ILIHPLDVLGLNRLILEYNV 244

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	92	7.5	473	4	US-09-134-001C-3405	Sequence 3405, Ap
2	87	7.1	815	4	US-09-134-001C-3614	Sequence 3614, Ap
3	86	7.0	653	2	US-08-713-928B-9	Sequence 9, Appl
4	86	7.0	653	4	US-09-191-171-5	Sequence 5, Appl
5	86	7.0	653	4	US-09-385-707-5	Sequence 5, Appl
6	86	7.0	653	4	US-09-439-923-2	Sequence 2, Appl
7	86	7.0	1140	4	US-09-220-081-2	Sequence 2, Appl
8	86	7.0	1140	4	US-09-677-575-2	Sequence 2, Appl
9	83	6.8	467	4	US-09-129-112-19	Sequence 19, Appl
10	79	6.4	741	4	US-09-001-984C-106	Sequence 106, App
11	79	6.4	6095	4	US-09-144-085-2	Sequence 2, Appl
12	78.5	6.4	271	4	US-09-318-794A-2	Sequence 2, Appl
13	78.5	6.4	271	4	US-09-318-793A-4	Sequence 4, Appl
14	77.5	6.3	3739	3	US-09-320-878-2	Sequence 2, Appl
15	77	6.3	554	4	US-08-462-467B-22	Sequence 22, Appl
16	77	6.3	983	2	US-08-164-292B-26	Sequence 26, Appl
17	77	6.3	983	3	US-08-845-623-26	Sequence 26, Appl
18	77	6.3	983	3	US-08-815-927-26	Sequence 26, Appl
19	77	6.3	983	4	US-09-103-330-26	Sequence 26, Appl
20	77	6.3	983	4	US-09-435-242-26	Sequence 26, Appl
21	77	6.3	1399	4	US-08-462-467B-14	Sequence 14, Appl
22	76.5	6.2	688	1	US-07-688-352C-28	Sequence 28, Appl
23	76.5	6.2	688	2	US-08-474-379C-28	Sequence 28, Appl
24	76.5	6.2	688	3	US-09-146-249A-28	Sequence 28, Appl
25	76.5	6.2	688	3	US-08-206-188B-28	Sequence 28, Appl
26	76	6.2	461	4	US-09-355-115-7	Sequence 7, Appl
27	76	6.2	1039	4	US-09-409-648-7	Sequence 7, Appl

28	76	6.2	1039	4	US-09-409-648-8	Sequence 8, Appl
29	76	6.2	1039	6	5196511-2	Patent No. 5196511
30	75.5	6.1	284	4	US-09-134-001C-2927	Sequence 2927, Ap
31	75	6.1	255	1	US-08-225-757B-2	Sequence 2, Appl
32	75	6.1	255	2	US-08-722-050-2	Sequence 2, Appl
33	75	6.1	415	4	US-09-182-450-1	Sequence 1, Appl
34	74	6.0	4536	4	US-09-180-422B-27	Sequence 27, Appl
35	73.5	6.0	529	4	US-08-887-534A-74	Sequence 74, Appl
36	73.5	6.0	3739	4	US-09-105-537-33	Sequence 33, Appl
37	73.5	6.0	11877	4	US-09-105-537-6	Sequence 6, Appl
38	73	5.9	993	1	US-08-444-792-2	Sequence 2, Appl
39	73	5.9	993	1	US-08-445-042-2	Sequence 2, Appl
40	72.5	5.9	480	4	US-09-004-838-53	Sequence 53, Appl
41	72.5	5.9	531	4	US-09-134-001C-4920	Sequence 4920, Ap
42	72.5	5.9	5087	4	US-09-144-085-1	Sequence 1, Appl
43	71.5	5.8	441	1	US-08-403-866-10	Sequence 10, Appl
44	71.5	5.8	2152	4	US-09-036-987A-3	Sequence 3, Appl
45	71.5	5.8	2152	4	US-09-370-700-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1  
US-09-134-001C-3405  
; Sequence 3405, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 3405  
; LENGTH: 473  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3405

Query Match	7.5%;	Score 92;	DB 4;	Length 473;
Best Local Similarity	24.9%;	Pred. NO. 0.098;		
Matches	46;	Conservative	22;	Mismatches 51; Indels 66; Gaps 9;
Qy	47	FDKAIYSSV--GLPDEEAAIVRSCAAASLMGMGTETVPPLRLQY----	DRRTLGA	DRLAA 100
Db	220	FEKQMTQPVKKGKKEGIEIVTEAMAKSA-----EETENGVKVITYEAKGEEQTEADYVLV		275
Qy	101	VVGAHSLYPNTE-----LLVIDAGTAITYERVSAGGIYLGGNISPGHLRL		145
Db	275	TVGRR---PNTDELGLEELGLKFA	DRGLLEVD-----KQSR	TSIENIFAIGDIVPGLPLA 327
Qy	145	FKALHFTGRPLIDPSGISPKIAEYGSSTEEAITAGVIHGLAGEIDRYIDDLHAK		EGRS 205
Db	328	HKA-----SYEGKVA	EAIDGQAAEVD-YI-----	GMP 354

RESULT 2  
US-09-134-001C-3614  
; Sequence 3614, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

;; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
;; FILE REFERENCE: GTC-007  
;; CURRENT APPLICATION NUMBER: US/09/134,001C  
;; CURRENT FILING DATE: 1998-08-13  
;; PRIOR APPLICATION NUMBER: US 60/064,964  
;; PRIOR FILING DATE: 1997-11-08  
;; PRIOR APPLICATION NUMBER: US 60/055,779  
;; PRIOR FILING DATE: 1997-08-14  
;; NUMBER OF SEQ ID NOS: 5674  
;; SEQ ID NO 3614  
;; LENGTH: 815  
;; TYPE: PRT  
;; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3614

Query Match 7.1%; Score 87; DB 4; Length 815;  
Best Local Similarity 20.9%; Pred. No. 0.82;  
Matches 69; Conservative 32; Mismatches 101; Indels 128; Gaps 12;

QY 4 NLIVDQNSACKVAFVRNNSIESISFELPGKAGQALSHLVAPHRFDKAIYSSVGLPDEEAE 63  
Db 117 NIVNTLGENAINDILTRPEILESVPSPKPKK-----KQIADQINANOSEK 163

QY 64 AIVR-----SCAAASLMMGTETPV-----PLRLQYDRRTLG---ADRLAAVVG--- 103  
Db 164 IMIRLHDLGFGPKLSMAIYQFYMGDTLNVLDKNPYQLVYDIKGIGFNKADQLARNVGIEP 223

QY 104 -----AHSLYPNT-----ELLVIDAGTAITYE----- 125  
Db 224 HSPERLKAALLFTLEBECIKQGHYLPRTIVTETQNLNEDIEKPIETEQLLEIDVLS 283

QY 126 -----RVSAGIYLGGNISPGHLHFRKALHFTGRPLIDPSGISPKIAE----- 170  
Db 284 EEKKLISEADQVSPISLYSELKSVQNLRYRIKT---NTSKLKEIEQSDLIHIGDIESQN 340

QY 171 ---YGSSTEEAITAGV-----IHGLAGEIDRYIDD-- 197  
Db 341 EVNYSASQKEALETAINSKIMLLTGGPGTGKTTVIKIGIVELYAEIHGLSLDYDDYEDDY 400

QY 198 ---LHAKGRSAVIL---TGGDANYLARI 221  
Db 401 PVVLAAPTGRASKRLHSTGLEAMTIHRLI 430

RESULT 3  
US-08-713-928B-9  
; Sequence 9, Application US/08713928B  
; Patent No. 5929304  
; GENERAL INFORMATION:  
; APPLICANT: RADIN, DAVID N.  
; APPLICANT: CRAMER, CAROLE L.  
; APPLICANT: OISHI, KAREN K.  
; APPLICANT: WEISSENBERN, DEBORAH L.  
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN  
; TITLE OF INVENTION: PLANT-BASED EXPRESSION SYSTEMS  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/713,928B  
; FILING DATE: 13-SEP-1996  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 60/003,737  
;; FILING DATE: 14-SEP-1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Coruzzi, Laura A.  
;; REGISTRATION NUMBER: 30,742  
;; REFERENCE/DOCKET NUMBER: 7956-0011-999  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (212) 790-9090  
;; TELEFAX: (212) 869-9741  
;; TELEX: 66141 PENNIE  
;; INFORMATION FOR SEQ ID NO: 9:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 653 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: unknown  
;; MOLECULE TYPE: peptide  
US-08-713-928B-9

Query Match 7.0%; Score 86; DB 2; Length 653;  
Best Local Similarity 20.7%; Pred. No. 0.75;  
Matches 53; Conservative 39; Mismatches 100; Indels 64; Gaps 10;

QY 25 ESISFELPGKAGQALSHLVAPHRFDKAIYSSVGLPDEE-----AEAIVRSCA 70  
Db 243 DCTNFTGEAGVRLDY--ISLHR--KGARSSISILEQEKVVAQEIRQLFPKFDATPIYNDE 299

QY 71 AASLMMGTETPVPLRLQYDRRTLGADRLAAVVGVAHSLYPNTELLVIDAGTAITYERSAE 130  
Db 300 ADPL-VGWSLPQWR-----ADVTYAAVVVKVIAHQHQLLLANTTSAPFYALLSND 349

QY 131 GIYL-----GGNISPGHLHFRKALHFTGRPLIDPSGISPKIAEYG 172  
Db 350 NAFLSYHPHPFAQRTLTARFQVNNTRPPHVQLLRKPVLTAMGLLALLDDEEQLWAEVSQAG 409

QY 173 SSTEAITAGVI---HGLAGEIDRY-----IDDLHAKGRSAVIL-----TGGDA 214  
Db 410 TVLDSNHTVGVLASAHRPQGPADAWRAAVLIYASDDTRAHPNRSVAVTLRLRGVPPGGL 469

QY 215 NYLARIIRSGILIHDP 230  
Db 470 VYVTRYLDNG-LCSPD 484

RESULT 4  
US-09-191-171-5  
; Sequence 5, Application US/09191171  
; Patent No. 6149909  
; Patent No. 6149909 6143294  
; GENERAL INFORMATION:  
; APPLICANT: SCOTT, Hamish S.  
; APPLICANT: ANSON, Donald S.  
; APPLICANT: ORSBORN, Annette M.  
; APPLICANT: NELSON, Paul V.  
; APPLICANT: CLEMENTS, Peter R.  
; APPLICANT: MORRIS, Charles P.  
; APPLICANT: HOPWOOD, John J.  
; TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDURONIDASE AND GENETIC  
; TITLE OF INVENTION: SEQUENCES ENCODING SAME  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SCULLY SCOTT MURPHY & PRESSER  
; STREET: 400 Garden City Plaza  
; CITY: Garden City  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 11530  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/191,171  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/494,104  
FILING DATE: 23-JUN-1995  
APPLICATION NUMBER: AU PK9490/91  
FILING DATE: 14-NOV-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: AU PCT/AU92/00611  
FILING DATE: 12-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/084,254  
FILING DATE: 07-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Digiglio, Frank S.  
REGISTRATION NUMBER: 31,346  
REFERENCE/DOCKET NUMBER: 8978Z  
TELEPHONE: 516-742-4343  
TELEFAX: 516-742-4366  
TELEX: 230 901 SANS UR  
SEQUENCE CHARACTERISTICS:  
LENGTH: 653 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-191-171-5

Query Match 7.0%; Score 86; DB 4; Length 653;  
Best Local Similarity 20.7%; Pred. No. 0.75;  
Matches 53; Conservative 39; Mismatches 100; Indels 64; Gaps 10;  
QY 25 ESISFLPGKAGQALSHLVAPHRFDKAIYSSVGLPDEE-----AEAIVRSCA 70  
Db 243 DGTNFFTEAGVRLDY-ISLHR--KGARSSISILEQEKVVAQQIRQLFPKFADTPIYNDE 299  
QY 71 AASLMMGTETPVPLRLQYDRRTLGADRLAAVVGAAHSLYPNTELLVIDAGTAITYERSVAE 130  
Db 300 ADPL-VGWSLPQPWR-----ADVTYAMVVKVIAQHONLLANTTSAPFYALLSND 349  
QY 131 GIYL-----GGNISPGHLRFLKALHLFTGRPLIDPSGISPKIAEYG 172  
Db 350 NAFLSYHPHPPAORTLTARFQVNNTRPPHVQLLRKPVLTAMGLLALLDEEQLWAEVSQAG 409  
QY 173 SSTEAITAGVI---HGLAGEIDRY-----IDDLHAKGRSAVIL-----TGGDA 214  
Db 410 TVLDSNHTVGVLASAHRPQGPADAWRAAVLIYASDDTRAHPNRSVAVTLRLRGVPPGPG 469  
QY 215 NYLARIIRSGILIHDP 230  
Db 470 VYVTRYLDNG-LCSPD 484  
RESULT 5  
US-09-385-707-5  
Sequence 5, Application US/09385707  
Patent No. 6238662  
GENERAL INFORMATION:  
APPLICANT: SCOTT, Hamish S.  
APPLICANT: ANSON, Donald S.  
APPLICANT: ORSBORN, Annette M.  
APPLICANT: NELSON, Paul V.  
APPLICANT: MORRIS, Peter R.  
APPLICANT: MORRIS, Charles P.  
APPLICANT: HOPWOOD, John J.  
TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDURONIDASE AND GENETIC  
NUMBER OF INVENTION: SEQUENCES ENCODING SAME  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SCULLY SCOTT MURPHY & PRESSER  
STREET: 400 Garden City plaza

CITY: Garden City  
STATE: New York  
COUNTRY: USA  
ZIP: 11530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/385,707  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/494,104  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Digiglio, Frank S.  
REGISTRATION NUMBER: 31,346  
REFERENCE/DOCKET NUMBER: 8978Z  
TELEPHONE: 516-742-4343  
TELEFAX: 516-742-4366  
TELEX: 230 901 SANS UR  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 653 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-385-707-5  
Query Match 7.0%; Score 86; DB 4; Length 653;  
Best Local Similarity 20.7%; Pred. No. 0.75;  
Matches 53; Conservative 39; Mismatches 100; Indels 64; Gaps 10;  
QY 25 ESISFLPGKAGQALSHLVAPHRFDKAIYSSVGLPDEE-----AEAIVRSCA 70  
Db 243 DGTNFFTEAGVRLDY-ISLHR--KGARSSISILEQEKVVAQQIRQLFPKFADTPIYNDE 299  
QY 71 AASLMMGTETPVPLRLQYDRRTLGADRLAAVVGAAHSLYPNTELLVIDAGTAITYERSVAE 130  
Db 300 ADPL-VGWSLPQPWR-----ADVTYAMVVKVIAQHONLLANTTSAPFYALLSND 349  
QY 131 GIYL-----GGNISPGHLRFLKALHLFTGRPLIDPSGISPKIAEYG 172  
Db 350 NAFLSYHPHPPAORTLTARFQVNNTRPPHVQLLRKPVLTAMGLLALLDEEQLWAEVSQAG 409  
QY 173 SSTEAITAGVI---HGLAGEIDRY-----IDDLHAKGRSAVIL-----TGGDA 214  
Db 410 TVLDSNHTVGVLASAHRPQGPADAWRAAVLIYASDDTRAHPNRSVAVTLRLRGVPPGPG 469  
QY 215 NYLARIIRSGILIHDP 230  
Db 470 VYVTRYLDNG-LCSPD 484  
RESULT 6  
US-09-439-923-2  
Sequence 2, Application US/09439923  
Patent No. 6426208  
GENERAL INFORMATION:  
APPLICANT: Emil D. Kakkis  
APPLICANT: Becky Tanamachi  
TITLE OF INVENTION: Recombinant Alpha-L-Iduronidase, Methods  
TITLE OF INVENTION: for Producing and Purifying the Same and Methods for  
TITLE OF INVENTION: Treating Diseases Caused by Deficiencies Thereof  
FILE REFERENCE: 08000051US00  
CURRENT APPLICATION NUMBER: US/09/439,923  
CURRENT FILING DATE: 1999-11-12  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: FastSEQ for Windows Version 3.0  
SEQ ID NO 2







```
RESULT 13
US-09-318-793A-4
; Sequence 4, Application US/09318793A
; Patent No. 6184007
; GENERAL INFORMATION:
; APPLICANT: Dusch, Nicole
; APPLICANT: Kalinowski, Jorn
; APPLICANT: Puhler, Alfred
; TITLE OF INVENTION: METHOD FOR THE FERMENTATIVE PRODUCTION OF D-PANTOTHENIC
; TITLE OF INVENTION: ACID BY ENHANCEMENT OF THE pand GENE IN MICROORGANISMS
; FILE REFERENCE: 21123/260204
; CURRENT APPLICATION NUMBER: US/09/318,793A
; CURRENT FILING DATE: 1999-05-26
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 271
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-318-793A-4

Query Match      6.4%; Score 78.5; DB 4; Length 271;
Best Local Similarity 20.2%; Pred. No. 1.3;
Matches 38; Conservative 35; Mismatches 80; Indels 35; Gaps 6;

QY 35 GQALSHLVAPHREDKAIYSSVGLPDEEAEAIVRSCAAASLMMGTETPVPLRLQYDRRTLG 94
Db 23 GQKVSVLTSYDALSARIFDEAGV-----DMLLVGDSAAENVLGRDITLSITL----- 69

QY 95 ADRLAAVVGAAHSLYPNTTELLVIDAGTAITYE-----RVSAEGIYLGNGISPLGH 143
Db 70 -DEMIVLAKAVTIATKRALVVVDLPFG-TYEVSPNQAVESAIRVMRETGAAAVKIEGGVE 127

QY 144 LRFKALHLFTGRPLIDPSGISPKIAEYGSSTEEAITAGVIHGLAGEIDRYIDDLHAKE- 202
Db 128 IAQTRIRIVDAGIPVVGHHIGYTPQ-----SEHSLGHHVVQGRGASSGKLIADARALEQ 180

QY 203 -GRSAVIL 209
Db 181 AGAFAVL 188

RESULT 14
US-09-320-878-2
; Sequence 2, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: McDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
```

```
; LENGTH: 3739
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-320-878-2

Query Match      6.3%; Score 77.5; DB 3; Length 3739;
Best Local Similarity 20.4%; Pred. No. 1e+02;
Matches 64; Conservative 31; Mismatches 110; Indels 109; Gaps 13;

QY 18 FVRNNSIESISFLPG-----KAGQAL-----SHLVAPHREDKAIYSSVGLP 58
Db 2456 WLADHAVAGTVLLPGTAFLAFRAGDQVGCDDLVEELTLDAPLVLPFRGAVRVQLSVGAS 2515

QY 59 DEEAFAIVRSCAAASLMMGTETPVPLRLQYDRRTLG-----ADRLAAVVGAAHSL-YPNTE 112
Db 2516 DESGRRTFG-----LYAHPEDAPGEAEWTRHATGVLAARADRTAPVADPEAWPPPGAE 2568

QY 113 LLVID-----AGTAITYERSAEGIYL----- 134
Db 2569 PVDVDGLYERFAANGYGYGPLFGQVGVWRRRGDEVFADVALPAEVAGAEGARFGLHPALL 2628

QY 135 -----GGNISPLHLRF--KALHLFT-----GRLPLIDPSGISPKIAEYGSSTE 176
Db 2629 DAAVQAAGAGAGAGAGTRLPFAWSGISLYAVGATALRVRLAPAGPDIVSVSAAD--SSGQ 2686

QY 177 EAITAGVI-----HGLAGEIDRYIDDLH-----AKEGRSAVILTGGDANYLARI 220
Db 2687 PVFAADSLTVLPVDPQAALAFSDFTLDALHLEWTANDGAAQALPGAVVLGGDADGLAAA 2746

QY 221 IRSG---ILIH PDL 231
Db 2747 LRAGGTEVLSFPDL 2760

RESULT 15
US-08-462-467B-22
; Sequence 22, Application US/08462467B
; Patent No. 6210899
; GENERAL INFORMATION:
; APPLICANT: Rosenbaum, Jan S
; TITLE OF INVENTION: The Use of a BMP Protein Receptor
; TITLE OF INVENTION: Complex for Screening Bone Metabolism Actives and Cells
; TITLE OF INVENTION: Co-Transfected With a Type II BMP Receptor and a Type I
; TITLE OF INVENTION: BMP Receptor
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Procter & Gamble Company
; STREET: 11810 East Miami River Road
; CITY: Ross
; STATE: OH
; COUNTRY: USA
; ZIP: 45061
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,467B
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hersko, Bart S.
; REGISTRATION NUMBER: 32,572
; REFERENCE/DOCKET NUMBER: 5474R
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (513) 627-0633
; TELEFAX: (513) 627-0260
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 554 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
```

; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
US-08-462-467B-22

Query Match 6.3%; Score 77; DB 4; Length 554;  
Best Local Similarity 28.6%; Pred. No. 5.9;  
Matches 40; Conservative 21; Mismatches 31; Indels 48; Gaps 11;  
QY 123 TY--ERVSAEGIYLG---NISPG-LHLRFKALHLFTGRLPLIDPS-----GISPKIAEY 171  
|| ||: ||: || | | || | | : : | : | : : |  
Db 439 TYRSERI-LEGLYLEGLGLEGLNASPGLTHRTYR-----ILEPRPRGLYGLSERLEY 487  
QY 172 GSSTEEAITAGVIHGLAGEIDRYIDDLHAKGRSAVILTGGDANYLARIIRS--GILIH 229  
: | | : : | : : : : : | | : | : | : |  
Db 488 SASPLE-----ILEGLGLNSERGLNSERSERGL-----YSERGLYSERGLYLEP 531  
QY 230 DL---VLLGLN-----RILE 241  
| | ||| |||  
b 532 RLELEVALGLNARGTHRILE 551

Search completed: June 24, 2003, 21:49:41  
Job time : 8.128 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:33:11 ; Search time 4.24384 Seconds  
(without alignments)  
2384.688 Million cell updates/sec

Title: US-09-813-453A-41  
Perfect score: 1229  
Sequence: 1 MSFNLIVDQGNACKVAFVR.....ILIHPLVLGLNRILEYNV 244

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues  
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description.
1	175.5	14.3	233	1	YACB_BACSU
2	163	13.3	267	1	BAF_BORPE
3	95.5	7.8	232	1	YPFH_ECOLI
4	95.5	7.8	468	1	DLDH_STAAM
5	95	7.7	332	1	ALF_TREPA
6	90.5	7.4	758	1	CHLD_TOBAC
7	88	7.2	417	1	U183_ARATH
8	88	7.2	555	1	FPRB_MYCLE
9	87.5	7.1	449	1	GSHR_BURCE
10	86	7.0	653	1	IDUA_HUMAN
11	85	6.9	389	1	AMHX_BACSU
12	84.5	6.9	941	1	GCSP_MYCTU
13	84	6.8	357	1	PYRD_MYCTU
14	84	6.8	612	1	LU1B_LYCPN
15	83.5	6.8	242	1	FABG_ACTAC
16	83.5	6.8	710	1	NASC_BACSU
17	83.5	6.8	1149	1	HDA6_MOUSE
18	83	6.8	360	1	BUK_ENTFA
19	83	6.8	422	1	PUR2_MYCLE
20	82.5	6.7	428	1	ORDL_RHIME
21	82	6.7	288	1	HEMK_MYCLE
22	82	6.7	405	1	CL30_MYCTU
23	82	6.7	816	1	SUS2_ORYSA
24	81.5	6.6	313	1	MRAW_SHEVI
25	81.5	6.6	612	1	ILVD_HAEIN
26	81	6.6	274	1	CCS_HUMAN
27	81	6.6	274	1	CCS_RAT
28	81	6.6	326	1	LEU3_ARCFU
29	81	6.6	339	1	TCMO_STRGA
30	81	6.6	430	1	TRPB_HALN1
31	81	6.6	561	1	MERA_PSEAE
32	80.5	6.6	255	1	YJIL_ECOLI
33	80.5	6.6	610	1	DNAK_METTE
					P37564 bacillus su
					Q45338 bordetella
					P76561 escherichia
					Q59822 staphylococ
					O83668 treponema p
					O24133 nicotiana t
					Q9sd33 arabidopsis
					O33064 mycobacteri
					P48639 burkholderi
					P35475 homo sapien
					P54983 bacillus su
					Q50601 mycobacteri
					O06236 mycobacteri
					O04974 lycopersico
					P70720 actinobacil
					P42434 bacillus su
					Q9z2v5 mus musculu
					Q9rps7 enterococcu
					Q50144 mycobacteri
					Q9z3s3 rhizobium m
					P45832 mycobacteri
					Q11062 mycobacteri
					P31924 oryza sativ
					Q9fln8 shewanella
					P44851 haemophilus
					O14618 homo sapien
					Q9jk72 rattus norv
					O29627 archaeoglob
					P39896 streptomyce
					Q9hsc0 halobacteri
					P00392 pseudomonas
					P39383 escherichia
					Q9uxr0 methanosarc

34	80.5	6.6	983	1	Y644_MYCLE	Q9ccm6 mycobacteri
35	80	6.5	331	1	SUCA_SCHPO	O13750 schizosacch
36	80	6.5	429	1	PUR2_VIBCH	Q9kv81 vibrio chol
37	80	6.5	600	1	RUBB_ARATH	P21240 arabidopsis
38	80	6.5	788	1	TRSL_HCMVA	P09695 human cytom
39	79.5	6.5	295	1	YUST_BACSU	O32186 bacillus su
40	79	6.4	492	1	FTSA_HELPY	O25629 helicobacte
41	79	6.4	494	1	VATB_PLAFA	Q25691 plasmodium
42	79	6.4	578	1	PTFB_RHOCA	P23387 rhodobacter
43	79	6.4	598	1	LEPA_HAEIN	P43729 haemophilus
44	79	6.4	598	1	LEPA_PASMU	P57806 pasteurella
45	79	6.4	741	1	MASZ_MYCTU	Q50596 mycobacteri

ALIGNMENTS

RESULT 1						
YACB_BACSU						
ID	YACB_BACSU	STANDARD;	PRT;	233	AA.	
AC	P37564;					
DT	01-OCT-1994 (Rel. 30, Created)					
DT	01-OCT-1994 (Rel. 30, Last sequence update)					
DT	15-JUN-2002 (Rel. 41, Last annotation update)					
DE	Hypothetical protein yacB.					
GN	YACB.					
OS	Bacillus subtilis.					
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.					
OX	NCBI_TaxID=1423;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=168;					
RX	MEDLINE=96051385; PubMed=7584024;					
RA	Ogasawara N., Nakai S., Yoshikawa H.;					
RT	"Systematic sequencing of the 180 kilobase region of the Bacillus					
RT	subtilis chromosome containing the replication origin.";					
RL	DNA Res. 1:1-14(1994).					
RN	[2]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=168;					
RX	MEDLINE=98044033; PubMed=9384377;					
RA	Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,					
RA	Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,					
RA	Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,					
RA	Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,					
RA	Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,					
RA	Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,					
RA	Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,					
RA	Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,					
RA	Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,					
RA	Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,					
RA	Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,					
RA	Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,					
RA	Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,					
RA	Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,					
RA	Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,					
RA	Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,					
RA	Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,					
RA	Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,					
RA	Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,					
RA	Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,					
RA	Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,					
RA	Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,					
RA	Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,					
RA	Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,					
RA	Tosato V., Uchiyama S., Vandenbol M., Vannier F., Weitzenecker T.,					
RA	Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,					
RA	Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,					
RA	Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;					
RT	"The complete genome sequence of the Gram-positive bacterium Bacillus					
RT	subtilis.";					
RL	Nature 390:249-256(1997).					
CC	-!- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.					







GN FBA OR FDA OR TP0662.  
OS Treponema pallidum.  
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
OX NCBI\_TaxID=160;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Nichols;  
RX MEDLINE=98332770; PubMed=9665876;  
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,  
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,  
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,  
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,  
RA McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,  
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,  
RA Venter J.C.;  
RT "Complete genome sequence of Treponema pallidum, the syphilis  
RT spirochete."  
RL Science 281:375-388(1998).  
CC -1- CATALYTIC ACTIVITY: D-fructose 1,6-bisphosphate - glycerone  
CC phosphate + D-glyceraldehyde 3-phosphate.  
CC -1- COFACTOR: ZINC (BY SIMILARITY).  
CC -1- PATHWAY: Glycolysis; sixth step.  
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO CLASS II FRUCTOSE-BISPHOSPHATE ALDOLASE  
CC FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE001240; AAC65635.1;  
DR TIGR; TP0662;  
DR InterPro; IPR000771; F\_bp\_aldolase.  
DR Pfam; PF01116; F\_bp\_aldolase; 1.  
DR ProDom; PD002376; F\_bp\_aldolase; 1.  
DR TIGRFAMS; TIGR00167; cbba; 1.  
DR PROSITE; PS00602; ALDOLASE\_CLASS\_II\_1; 1.  
DR PROSITE; PS00806; ALDOLASE\_CLASS\_II\_2; FALSE\_NEG.  
KW Lyase; Glycolysis; Zinc; Complete proteome.  
FT METAL 91 91 ZINC (BY SIMILARITY).  
FT METAL 94 94 ZINC (BY SIMILARITY).  
SQ SEQUENCE 332 AA; 36189 MW; B6D8D56BF99B52EA CRC64;  
Query Match 7.7%; Score 95; DB 1; Length 332;  
Best Local Similarity 21.7%; Pred. No. 1.1;  
Matches 49; Conservative 33; Mismatches 76; Indels 68; Gaps 8;  
QY 47 FDKAIYSSVGLPD-----EEAEIVRSCAAASLMGMGTETPVPLRLQYDRRTLGLADRLAA 100  
Db 16 FAKAVKGGYAIAPAYNFNNLEQLQAIQAC-----VETRSPVILQVSSGARK----- 61  
QY 101 VVGAHSLYPNTLLVIDAGTAITYERVSAEGTYLGGNISPLHL-----R 145  
Db 62 -----YANATLLRNMGAVEYAH-----ELGVDIPIVLHDHGDLSLELCIDCIESG 108  
QY 146 FKALHFTGRPLIDPSGISPKIAEYGSSTEEAIT-----AGVIHGLAGEIDRY--- 194  
Db 109 FSSVMIDGSAIPYDENVALSRKVCEYAHARADYVTVEGELGVLGAGVEDDVVAEKSHYTMP 168  
QY 195 --IDDLHAKEGRSAVILTGGDANYLARI-----RSGILIHPLD 231  
Db 169 DEVEDFVKTKGVDSLAISIGTSHGRAKFTPEQCTRNADGVLIPPL 214  
RESULT 6  
CHLD\_TOBAC  
ID CHLD\_TOBAC STANDARD; PRT; 758 AA.  
AC O24133;  
DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Magnesium-chelatase subunit chLD, chloroplast precursor (Mg-  
DE protoporphyrin IX chelatase) (Mg-chelatase subunit D).  
GN CHLD.  
OS Nicotiana tabacum (Common tobacco).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.  
OX NCBI\_TaxID=4097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98079235; PubMed=9418040;  
RA Papenbrock J., Graefe S., Kruse E., Haenel F., Grimm B.;  
RT "Mg-chelatase of tobacco: identification of a Chl D cDNA sequence  
RT encoding a third subunit, analysis of the interaction of the three  
RT subunits with the yeast two-hybrid system, and reconstitution of the  
RT enzyme activity by co-expression of recombinant CHL D, CHL H and CHL  
RT I.";  
RL Plant J. 12:981-990(1997).  
CC -1- FUNCTION: Involved in chlorophyll biosynthesis; introduces a  
CC magnesium ion into protoporphyrin IX to yield Mg-protoporphyrin  
CC IX.  
CC -1- PATHWAY: Chlorophyll biosynthesis.  
CC -1- SUBCELLULAR LOCATION: Chloroplast stroma.  
CC -1- SIMILARITY: BELONGS TO THE MG-CHELATASE SUBUNITS D/I FAMILY.  
CC -1- SIMILARITY: CONTAINS 1 VWFA DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Y10022; CAA71128.1; --  
DR InterPro; IPR000523; Mg\_chelatase\_chII.  
DR InterPro; IPR002035; VWF\_A.  
DR Pfam; PF01078; Mg\_chelatase; 1.  
DR SMART; SM00327; VWF; 1.  
DR PROSITE; PS0234; VWFA; 1.  
KW Photosynthesis; Chlorophyll biosynthesis; Chloroplast;  
KW Transit peptide.  
FT TRANSIT 1 62 CHLOROPLAST (POTENTIAL).  
FT CHAIN 63 758 MAGNESIUM-CHELATASE SUBUNIT CHLD.  
FT DOMAIN 556 752 VWFA.  
FT DOMAIN 399 445 GLU/PRO-RICH.  
FT DOMAIN 409 416 POLY-PRO.  
FT DOMAIN 427 434 POLY-GLU.  
SQ SEQUENCE 758 AA; 83012 MW; ED531E85D5FAFEF8 CRC64;  
Query Match 7.4%; Score 90.5; DB 1; Length 758;  
Best Local Similarity 22.8%; Pred. No. 7.1;  
Matches 49; Conservative 32; Mismatches 59; Indels 75; Gaps 11;  
QY 83 PLRLQYDRRTLGLADRLAAVVGAAHSLYPNTLLVID---AGFAITYERVSAEGTYLGGNIS 139  
Db 70 PEKISFGRQYF---PLAAVIGQDAIKTALLGLAIDREIGGTAICGKRGTAKTLL-----MA 121  
QY 140 PGLHLRFKALHFTGRPLIDPS-----GISPKIAEYGS----- 173  
Db 122 RGLHAILPPIEVVVGSMANADPNCPDEWEDGLADR-AEYSDGNIKTOIVKSPFVQIPLG 180  
QY 174 -----STEEAITAGVI---HGLAGEIDR---YDDLHAKEGRSAVILTGGDANYL 217  
Db 181 VTEDRLIGSVDVESVKSQTTVPQGLLAEAHRGVLYDEIN-----LLDEGISNLL 232  
QY 218 ARIIRSGILI-----HPDLVLLGLNRILEYN 243  
Db 233 LNVTEGVNIVEREGISFRHPCKPLL-----IATYN 263





```
FT METAL          9          IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT METAL        15          IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT METAL        19          IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT METAL        46          IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT METAL        49          IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT METAL        52          IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT METAL        56          IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
SQ SEQUENCE    555 AA; 59712 MW; 43C7292E6A2DFED9 CRC64;

Query Match          7.2%; Score 88; DB 1; Length 555;
Best Local Similarity 25.4%; Pred. No. 7.8;
Matches 51; Conservative 24; Mismatches 58; Indels 68; Gaps 12;

QY 29 FLPGKAGQALSH--LVAPHRFDKAIYSSVGLPDEEAEAIVRSCAAASLMMGTETPVPLRL 86
   || : ||| ||| ||| : ||| ||| : ||| ||| : ||| ||| : ||| ||| : |||
Db 182 FLNVEIGHLSHDELLAHHH---AVLYAVGAPDDR-----RL 215

QY 87 QYDRR---TLGADRLAAVVGASHLYPNTELLVIDAGTAITYERSAEGIYLGNIISPG 142
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 216 NIDGMGIPGTGATELVAVINAH---PDFAYLPVD---LSHERVVVIG---NGNVA--- 262

QY 143 HLRFKALHFTGRLPLIDPSGIS-PKIAEY-----GSSTEEAITA---GVIHGLAGEID 192
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 263 -----LDVARLLTADPDNLARTDISEFALHVLGSAVREVVVAARRGPAHSFAF--- 310

QY 193 RYIDDLHAKEGRSAVILTGGD 213
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 311 -TLPELIGLKATSEVVLDAGD 330
```

```
RESULT 9
GSHR_BURCE      STANDARD;      PRT;      449 AA.
AC P48639;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Glutathione reductase (EC 1.6.4.2) (GR) (Grase).
GN GOR.
OS Burkholderia cepacia (Pseudomonas cepacia).
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC Burkholderia.
OX NCBI_TaxID=292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AC1100;
RX MEDLINE=95266809; PubMed=7538273;
RA Daubaras D.L., Hersberger C.D., Kitano K., Chakrabarty A.M.;
RT "Sequence analysis of a gene cluster involved in metabolism of 2,4,5-
  trichlorophenoxyacetic acid by Burkholderia cepacia AC1100.";
Appl. Environ. Microbiol. 61:1279-1289(1995).
CC -|- CATALYTIC ACTIVITY: NADPH + oxidized glutathione = NADP(+) + 2
  glutathione.
CC -|- COFACTOR: FAD.
CC -|- PATHWAY: 2,4,5-trichlorophenoxyacetic acid degradation.
CC -|- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: Cytoplasmic.
CC -|- MISCELLANEOUS: THE ACTIVE SITE IS A REDOX-ACTIVE DISULFIDE BOND.
CC -|- SIMILARITY: BELONGS TO THE PYRIDINE NUCLEOTIDE-DISULFIDE
  OXIDOREDUCTASES CLASS-I.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U19883; AAC43334.1; -
CC HSSP; P28593; 1AOG.
DR InterPro; IPR001327; FAD_pyr_redox.
DR InterPro; IPR001100; Pyr_redox.
```

```
DR InterPro; IPR004099; pyr_redox_dim.
DR Pfam; PF00707; pyr_redox; 1.
DR Pfam; PF02852; pyr_redox_dim; 1.
DR PRINTS; PR00368; FADPNR.
DR PRINTS; PR00411; PNDRDTASEI.
DR ProDom; PD000139; FAD_pyr_redox; 1.
DR PROSITE; PS00076; PYRIDINE_REDOX_1; 1.
KW Redox-active center; Oxidoreductase; Flavoprotein; FAD; NADP.
FT NP_BIND      7 37          FAD (ADP PART) (PROBABLE).
FT DISULFID     43 48          REDOX-ACTIVE (BY SIMILARITY).
FT NP_BIND     292 302          FAD (FLAVIN PART) (BY SIMILARITY).
FT ACT_SITE    435 435          BY SIMILARITY.
SQ SEQUENCE    449 AA; 47541 MW; 402FCC6E7A8D6720 CRC64;

Query Match          7.1%; Score 87.5; DB 1; Length 449;
Best Local Similarity 22.7%; Pred. No. 6.7;
Matches 60; Conservative 35; Mismatches 82; Indels 87; Gaps 14;

QY 6 IVDQGSACKVAFVRNNSIESISFLPGKAGQALSHLVAPHR-----FDKAIYSSVGLP 58
   | : | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 172 IIGGGYIACEFAGIFNG-----LGR---HVVQLHRGSQVLRGFDDELREHLG-- 215

QY 59 DE-----EAEAIVRSCAAASLMMGTETPVPLRQYDRRTLGAADRLAAVVG 106
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 216 DELKKSIGIDLRLGVVAVVERQSGALSVQLTTG-----DAMEVDVMAATG 261

QY 107 LYPNT---ELLVIDAGT---AI---TYERSAEGIYLGNIISPGHLRFLKALH----- 150
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 262 RLPNTWGLGLETVDVGLDQNGAIKVDSEYRTSSPGIYAVGVDVNTNLTTPVAIHGHAFA 321

QY 151 --LFTGRLPLIDPSGI-----SPKIAEYSGSTEEAITAGVIHGLAGEIDRYID-DLHAK 201
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 322 DTVEGGKALPTEHENVPFAVFSQPQAASVGLSEAQA-----RDYSNVEIYGS 369

QY 202 EGR-SAVILTTGGDANYLARIIRSG 224
   | : | | | | | : | : | : | : | : | : | : | : | : | : | : |
Db 370 AFRPMRAALSGRDEKALVKLVVNG 393
```

```
RESULT 10
IDUA_HUMAN      STANDARD;      PRT;      653 AA.
AC P35475;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alpha-L-iduronidase precursor (EC 3.2.1.76).
GN IDUA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=92052158; PubMed=1946389;
RA Scott H.S., Anson D.S., Orsborn A.M., Nelson P.V., Clements P.R.,
RA Morris C.P., Hopwood J.J.;
RT "Human alpha-L-iduronidase: cDNA isolation and expression.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:9695-9699(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92372031; PubMed=1505961;
RA Scott H.S., Guo X.H., Hopwood J.J., Morris C.P.;
RT "Structure and sequence of the human alpha-L-iduronidase gene.";
RL Genomics 13:1311-1313(1992).
RN [3]
RP REVIEW ON VARIANTS.
RX MEDLINE=96287378; PubMed=8680403;
RA Scott H.S., Bunge S., Gal A., Clarke L.A., Morris C.P., Hopwood J.J.;
RT "Molecular genetics of mucopolysaccharidosis type I: diagnostic,
  clinical, and biological implications.";
RL Hum. Mutat. 6:288-302(1995).
```

RN [4] VARIANTS MPS-I THR-75.  
RP MEDLINE-94290500; PubMed=8019563;  
RX Clarke L.A., Nelson P.V., Warrington C.L., Morris C.P., Hopwood J.J.,  
RA Scott H.S.;  
RT "Mutation analysis of 19 North American mucopolysaccharidosis type I  
RT patients: identification of two additional frequent mutations.";  
RL Hum. Mutat. 3:275-282(1994).  
RN [5]  
RP VARIANTS MPS-I PRO-82.  
RX MEDLINE-94004880; PubMed=8401515;  
RA Clark L.A., Scott H.S.;  
RT "Two novel mutations causing mucopolysaccharidosis type I detected by  
RT single strand conformational analysis of the alpha-L-iduronidase  
RT gene.";  
RL Hum. Mol. Genet. 2:1311-1312(1993).  
RN [6]  
RP VARIANTS MPS-I GLN-89.  
RX MEDLINE-94027086; PubMed=8213840;  
RA Scott H.S., Litjens T., Nelson P.V., Thompson P.R., Brooks D.A.,  
RP Hopwood J.J., Morris C.P.;  
RT "Identification of mutations in the alpha-L-iduronidase gene (IDUA)  
RT that cause Hurler and Scheie syndromes.";  
RL Am. J. Hum. Genet. 53:973-986(1993).  
RN [7]  
RP VARIANTS MPS-I PRO-366 AND ARG-409.  
RX MEDLINE-93318833; PubMed=8328452;  
RA Bach G., Moskowitz S.M., Tieu P.T., Matynia A., Neufeld E.F.;  
RT "Molecular analysis of Hurler syndrome in Druze and Muslim Arab  
RT patients in Israel; multiple allelic mutations of the IDUA gene in a  
RT small geographic area.";  
RL Am. J. Hum. Genet. 53:330-338(1993).  
RN [8]  
RP VARIANTS MPS-I ARG-533.  
RX MEDLINE-93250829; PubMed=1301941;  
RA Scott H.S., Litjens T., Nelson P.V., Brooks D.A., Hopwood J.J.,  
RA Morris C.P.;  
RT "Alpha-L-iduronidase mutations (Q70X and P533R) associate with a  
RT severe Hurler phenotype.";  
RL Hum. Mutat. 1:333-339(1992).  
RN [9]  
RP VARIANTS MPS-I D-51; T-75; P-218; P-327; P-489 AND 16-S--A-19 DEL.  
RX MEDLINE-95038736; PubMed=7951228;  
RA Bunge S., Kleijer W.J., Steglich C., Beck M., Zuther C., Morris C.P.,  
RA Schwinger E., Hopwood J.J., Scott H.S., Gal A.;  
RT "Mucopolysaccharidosis type I: identification of 8 novel mutations  
RT and determination of the frequency of the two common  
RT alpha-L-iduronidase mutations (W402X and Q70X) among European  
RT patients.";  
RL Hum. Mol. Genet. 3:861-866(1994).  
RN [10]  
RP VARIANTS HIS-33.  
RX MEDLINE-93138632; PubMed=1362562;  
RA Scott H.S., Litjens T., Hopwood J.J., Morris C.P.;  
RT "PCR detection of two RFLPs in exon I of the alpha-L-iduronidase  
RT (IDUA) gene.";  
RL Hum. Genet. 90:327-327(1992).  
RN [11]  
RP VARIANTS THR-361.  
RX MEDLINE-94061048; PubMed=8242073;  
RA Scott H.S., Nelson P.V., Litjens T., Hopwood J.J., Morris C.P.;  
RT "Multiple polymorphisms within the alpha-L-iduronidase gene (IDUA):  
RT implications for a role in modification of MPS-I disease phenotype.";  
RL Hum. Mol. Genet. 2:1471-1473(1993).  
RN [12]  
RP VARIANTS MPS-I PRO-490; PRO-492 AND LEU-496.  
RX MEDLINE-96055518; PubMed=7550232;  
RA Tieu P.T., Bach G., Matynia A., Hwang M., Neufeld E.F.;  
RT "Four novel mutations underlying mild or intermediate forms of  
RT alpha-L-iduronidase deficiency (MPS IS and MPS IH/S).";  
RL Hum. Mutat. 6:55-59(1995).  
RN [13]  
RP VARIANTS MPS-I W-89; 349-D-N-350 DEL; H-383; T-504 AND R-626.

RX MEDLINE-96055526; PubMed=7550242;  
RA Bunge S., Kleijer W.J., Steglich C., Beck M., Schwinger E., Gal A.;  
RT "Mucopolysaccharidosis type I: identification of 13 novel mutations  
RT of the alpha-L-iduronidase gene.";  
RL Hum. Mutat. 6:91-94(1995).  
RN [14]  
RP VARIANTS MPS-I ARG-388.  
RA Barthclomew D.W., McClellan J.M.;  
RT "A novel missense mutation in the human IDUA gene associated with a  
RT severe Hurler's phenotype.";  
RL Hum. Mutat. 12:291-291(1998).  
CC -I- CATALYTIC ACTIVITY: Hydrolysis of alpha-L-iduronosidic linkages in  
CC desulfated dermatan.  
CC -I- SUBUNIT: MONOMER (PROBABLE).  
CC -I- SUBCELLULAR LOCATION: Lysosomal.  
CC -I- TISSUE SPECIFICITY: FOUND UBIQUITOUSLY.  
CC -I- DISEASE: DEFECTS IN IDUA ARE THE CAUSE OF MUCOPOLYSACCHARIDOSIS  
CC TYPE I (MPS-I). A DISEASE CHARACTERIZED BY THE ACCUMULATION OF THE  
CC GLYCOSAMINOGLYCAN HEPARAN SULFATE AND DERMATAN SULFATE. PATIENTS  
CC WITH SEVERE MPS-I (HURLER SYNDROME) USUALLY PRESENT WITHIN THE  
CC FIRST YEAR OF LIFE A COMBINATION OF HEPATOSPLENOMEGALY, SKELETAL  
CC DEFORMITIES, CORNEAL CLOUDING AND SEVERE MENTAL RETARDATION.  
CC OBSTRUCTIVE AIRWAYS DISEASE, RESPIRATORY INFECTION AND CARDIAC  
CC COMPLICATIONS USUALLY RESULTS IN DEATH BEFORE 10 YEARS OF AGE.  
CC PATIENTS WITH MILD MPS-I (SCHEIE SYNDROME) MAY HAVE LITTLE OR NO  
CC NEUROLOGICAL INVOLVEMENT, NORMAL STATURE AND LIFE SPAN, BUT ARE  
CC CHARACTERIZED BY THE DEVELOPMENT OF JOINTS STIFFNESS, MILD  
CC HEPATOSPLENOMEGALY, AORTIC VALVE DISEASE AND CORNEAL CLOUDING. THE  
CC INTERMEDIATE FORM (HURLER/SCHEIE) MAY BE CHARACTERIZED BY  
CC RELATIVELY LITTLE NEUROLOGICAL INVOLVEMENT, BUT MOST OF THE  
CC SOMATIC SYMPTOMS DESCRIBED FOR SEVERE MPS-I DEVELOP IN THE EARLY  
CC TO MID-TEENS, CAUSING CONSIDERABLE LOSS OF MOBILITY.  
CC -I- SIMILARITY: BELONGS TO FAMILY 39 OF GLYCOSYL HYDROLASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M74715; AAA81589.1; -  
CC EMBL; M95740; AAA51698.1; -  
CC EMBL; M95739; AAA51698.1; JOINED.  
CC Genew; HGNC:5391; IDUA.  
CC MIM; 252800; -  
CC InterPro; IPR000514; Glyco\_hydro\_39.  
CC Pfam; PF01229; Glyco\_hydro\_39; 1.  
CC PRINTS; PR00745; GLHYDRLASE39.  
CC PROSITE; PS01027; GLYCOSYL\_HYDROL\_F39; 1.  
KW Hydrolase; Glycosidase; Lysosome; Signal; Disease mutation;  
KW Polymorphism; Mucopolysaccharidosis.  
FT SIGNAL 1 27 POTENTIAL.  
FT CHAIN 28 653 ALPHA-L-IDURONIDASE.  
FT ACT\_SITE 182 182 PROTON DONOR (POTENTIAL).  
FT ACT\_SITE 299 299 NUCLEOPHILE (POTENTIAL).  
FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 336 336 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 372 372 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 415 415 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 451 451 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT VARIANT 16 19 MISSING (IN MPS-I; HURLER).  
FT /FTid=VAR\_003349.  
FT O -> H.  
FT VARIANT 33 33 /FTid=VAR\_003350.  
FT G -> D (IN MPS-I; HURLER).  
FT VARIANT 51 51 /FTid=VAR\_003351.  
FT A -> T (IN MPS-I; HURLER).  
FT VARIANT 75 75 /FTid=VAR\_003352.  
FT H -> P (IN MPS-I; HURLER/SCHEIE).  
FT /FTid=VAR\_003353.









RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains.";  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
CC -|- CATALYTIC ACTIVITY: (S)-dihydroorotate + O(2) -> orotate +  
CC H(2)O(2).  
CC -|- COFACTOR: FMN (BY SIMILARITY).  
CC -|- PATHWAY: Pyrimidine biosynthesis; fourth step.  
CC -|- SUBUNIT: HOMODIMER (BY SIMILARITY).  
CC -|- SUBCELLULAR LOCATION: INNER SIDE OF THE MEMBRANE (BY SIMILARITY).  
CC -|- SIMILARITY: BELONGS TO THE DIHYDROOROTATE DEHYDROGENASE FAMILY.  
CC SUBFAMILY 2.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z95388; CAB08654.1; --  
CC EMBL; AE007067; AAK46481.1; --  
CC TIGR; MT2197; --  
CC TubercuList; RV2139; --  
CC InterPro; IPR001295; DHO\_dh.  
CC InterPro; IPR003009; FMN\_enzyme.  
CC Pfam; PF01180; DHodehase; 1.  
CC TIGRFAMS; TIGR01036; pyrd\_sub2; 1.  
CC PROSITE; PS00911; DHODEHASE\_1; 1.  
CC PROSITE; PS00912; DHODEHASE\_2; 1.  
KW Pyrimidine biosynthesis; Oxidoreductase; Flavoprotein; FMN;  
KW Complete proteome.  
FT NP\_BIND 286 294 FMN (POTENTIAL).  
SQ SEQUENCE 357 AA; 37998 MW; 3D9D107DD9B4FCB6 CRC64;  
  
Query Match 6.8%; Score 84; DB 1; Length 357;  
Best Local Similarity 26.1%; Pred. No. 9.8;  
Matches 64; Conservative 27; Mismatches 100; Indels 54; Gaps 12;  
  
QY 1 MSFNLIVDQGNACKVAFVRNNSIESISFLPGK-----AGQALSHLVAPHRFDKAIYS-- 53  
Db 117 MGFN---NHGARALAIRLARHREPIPIGVNIGTKTKTPAGDAVNDYRASARWVGPLASYL 173  
  
QY 54 -----SVGLPDEEAIAIVRSCAAASLMMGTTETPVPLRLQYDRRTLGLADRLAAVVGAH 105  
Db 174 VNVSSPNTPLRLDQAVESLRPLISA-VRAETSTPVLVKIAPDLSDSLDDIADLAVEL 232  
  
106 SLYPNTLLVIDAGTAITYERSAEGIYL-----GGNISPLHLR----FKALHLFT 153  
Db 233 DL-----AGIVATNTTVSRDGLTTPGVDRLPGPGISGPPLAQRAVQVLRRLYDRV 282  
  
QY 154 G-RLPLIDPSGISPKIAEYGSSTEEAITAG--VIHGLAGEI---DRYIDDLHAKEGRSVAV 207  
Db 283 GDRIALISVGGI-----ETADDAWERITAGASLLQGYTGFTYGGERWAKDIH--EGIARR 335  
  
QY 208 ILTGG 212  
Db 336 LHDGG 340

RESULT 14  
LULB\_LYCPN STANDARD; PRT; 612 AA.  
AC 004974;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE 2-isopropylmalate synthase B (EC 4.1.3.12) (Alpha-isopropylmalate  
synthase B) (Alpha-IPM synthetase B).  
GN IPMSB.

OS Lycopersicon pennellii (Tomato).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
OX NCBI\_TaxID=28526;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Leaf;  
RA Wei T., Maita D., Steffens J.C.;  
RT "Cloning of two L. pennellii 2-isopropylmalate synthase cDNA and  
RT their functional expression in yeast.";  
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
CC -|- FUNCTION: Catalyzes the condensation of the acetyl group of  
CC acetyl-CoA with 3-methyl-2-oxobutanoate (2-oxoisovalerate) to form  
CC 3-carboxy-3-hydroxy-4-methylpentanoate (2-isopropylmalate).  
CC -|- CATALYTIC ACTIVITY: 3-carboxy-3-hydroxy-4-methylpentanoate + CoA ->  
CC acetyl-CoA + 3-methyl-2-oxobutanoate + H(2)O.  
CC -|- PATHWAY: Leucine biosynthesis; first step.  
CC -|- SIMILARITY: BELONGS TO THE ALPHA-IPM SYNTHETASE / HOMOCITRATE  
CC SYNTHASE FAMILY. LEUA 1 SUBFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF004166; AAB61599.1; --  
CC InterPro; IPR002034; AIPM/Hcit\_synth.  
CC InterPro; IPR000891; HMGL-like.  
CC Pfam; PF00682; HMGL-like; 1.  
CC TIGRFAMS; TIGR00973; leua\_bact; 1.  
CC PROSITE; PS00815; AIPM\_HOMOCIT\_SYNTH\_1; 1.  
CC PROSITE; PS00816; AIPM\_HOMOCIT\_SYNTH\_2; FALSE\_NEG.  
KW Leucine biosynthesis; Lyase.  
SQ SEQUENCE 612 AA; 66535 MW; 2EC5AEC9039791B CRC64;  
  
Query Match 6.8%; Score 84; DB 1; Length 612;  
Best Local Similarity 23.3%; Pred. No. 19;  
Matches 60; Conservative 31; Mismatches 100; Indels 66; Gaps 12;  
  
QY 1 MSFNLIVDQGNACKVAFVRNNSIESISFLPGKAGQA----LSHLVAPHRFDKAIYSSVG 56  
Db 187 MSRDQVVEKARS--MVAYARSIGCEDVEFSPEDAGRSDFLYHILG--EVIKAGATTLN 242  
  
QY 57 LPDEEAIAIVRSCAAASLMMGTTETP-----VPLRLQYDRRTLGLADRLA-AVVGASLYP 109  
Db 243 IPDTVGYTVPEEFGQLIAKIKANTPGVEDVIISTHCQNDLGLSTANTLAGACAGARQLEV 302  
  
QY 110 NTELLVIDAGTAITYERSVSA-----EGIYLGNIS-----PGLHLR-FK 147  
Db 303 TINGIGERAGNASLEEVVMALKRCGEQVGLGLYTGINTHILMSSKMWEGISGLHVQPHK 362  
  
QY 148 ALHLFTGRLPLIDPSG-----ISPKIAEYGSSTEEAITAGVIHG----- 186  
Db 363 AI---VGANAFVHESGIHQDGMKHKDKTYEISPIDGLNRANESGIVFGKLSGVMCKP 419  
  
QY 187 ----LAGEID-RYIDDL 198  
Db 420 KMELGYEIEGKELDDL 436

RESULT 15  
FABG\_ACTAC STANDARD; PRT; 242 AA.  
ID FABG\_ACTAC  
AC P70720;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE 3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100) (3-ketoacyl-  
DE acyl carrier protein reductase).

Search completed: June 24, 2003, 21:51:34  
Job time : 6.24384 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:26 ; Search time 25.051 Seconds  
(without alignments)  
1297.879 Million cell updates/sec

Title: US-09-813-453A-41  
Perfect score: 1229  
Sequence: 1 MSFNLIYDQNSACKVAFVR.....ILIHDPDLVLLGLNRILEYNV 244

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

arched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002:\*  
1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1229	100.0	244	AAU91168	Pantothenate kinas
2	224	18.2	258	AAU91153	Rhodobacter capsul
3	221	18.0	257	AAU91174	Pantothenate kinas
4	216	17.6	259	ABB47661	Listeria monocytog
5	210.5	17.1	258	AAU91172	Pantothenate kinas
6	208	16.9	255	AAU91154	Geobacter sulfurre
7	207.5	16.9	262	AAU91170	Pantothenate kinas
8	207	16.8	258	AAU01243	B. subtilis novel
9	207	16.8	258	AAU91149	Bacillus subtilis
10	205.5	16.7	265	AAU91151	Streptomyces coeli

11	201.5	16.4	256	23	AAU91175	Pantothenate kinas
12	198	16.1	260	23	AAU91173	Pantothenate kinas
13	197.5	16.1	254	23	AAU91171	Pantothenate kinas
14	195.5	15.9	248	23	AAU91164	Pantothenate kinas
15	184	15.0	219	23	AAU91176	Pantothenate kinas
16	183.5	14.9	241	23	AAU91179	Pantothenate kinas
17	181.5	14.8	242	23	AAU91180	Pantothenate kinas
18	181	14.7	249	23	AAU91178	Pantothenate kinas
19	178	14.5	249	23	AAU91182	Pantothenate kinas
20	175.5	14.3	233	23	AAU91163	Pantothenate kinas
21	174	14.2	212	23	AAU91177	Pantothenate kinas
22	172.5	14.0	273	23	AAU91157	Treponema pallidum
23	167.5	13.6	250	23	AAU91150	Clostridium acetob
24	166	13.5	455	20	AAU91156	Neisseria gonorrhoe
25	166	13.5	455	21	AAU91156	Neisseria gonorrhoe
26	166	13.5	460	23	AAU91167	Pantothenate kinas
27	166	13.5	592	20	AAU91167	Neisseria gonorrhoe
28	166	13.5	592	21	AAU91167	Neisseria gonorrhoe
29	163	13.3	267	23	AAU91162	Bordetella pertussis
30	162.5	13.2	389	21	AAU91162	Neisseria meningit
31	162	13.2	246	23	AAU91156	Thermotoga maritim
32	161	13.1	592	20	AAU91156	Neisseria meningit
33	161	13.1	592	21	AAU91156	Neisseria meningit
34	161	13.1	592	23	AAU91169	Neisseria meningit
35	156	12.7	455	21	AAU91169	Pantothenate kinas
36	156	12.7	592	20	AAU91166	Neisseria meningit
37	156	12.7	592	21	AAU91166	Neisseria meningit
38	156	12.7	592	23	AAU91166	Pantothenate kinas
39	153	12.4	272	22	AAU91166	Mycobacterium tube
40	153	12.4	272	23	AAU91166	Mycobacterium tube
41	152	12.4	257	23	AAU91160	Synechocystis pant
42	135	11.0	262	23	AAU91158	Borrelia burgdorfe
43	132.5	10.8	229	23	AAU91159	Aquifex aeolicus p
44	131	10.7	262	23	AAU91155	Deinococcus radiop
45	129.5	10.5	223	19	AAU91155	H. pylori GHPO 344

ALIGNMENTS

RESULT 1  
AAU91168  
ID AAU91168 standard; Protein; 244 AA.  
XX  
AC AAU91168;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Pantothenate kinase (Coax) #6.  
XX  
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX  
OS Porphyromonas gingivalis.  
XX  
PN WO200216601-A2.  
XX  
PD 28-FEB-2002.  
XX  
PF 24-AUG-2001; 2001WO-US26531.  
XX  
PR 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX  
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX  
PI Yocum RR, Patterson TA;  
XX  
DR WPI; 2002-269358/31.  
DR N-PSDB; ABK54189.  
XX  
PT Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein

PT with test compound and identifying inhibitor of the Coax protein  
XX  
PS Claim 8; Page 94-95; 128pp; English.  
XX  
CC The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
XX described in the invention.  
SQ Sequence 244 AA;

Query Match 100.0%; Score 1229; DB 23; Length 244;  
Best Local Similarity 100.0%; Pred. No. 9.4e-119;  
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 MSFNLIVDQGSACKVAFVRNNSIESISFLPGKAGQALSHLVAPHRFDKAIYSSVGLPDE 60  
1 MSFNLIVDQGSACKVAFVRNNSIESISFLPGKAGQALSHLVAPHRFDKAIYSSVGLPDE 60  
61 EAEIVRSCAAASLMMGTETPVPLRLQYDRRTLGAADRLAAVVGAAHSLYPNTELLVIDAGT 120  
61 EAEIVRSCAAASLMMGTETPVPLRLQYDRRTLGAADRLAAVVGAAHSLYPNTELLVIDAGT 120  
121 AITYERSAEGIYLGNGISPLGLHFRKALHFTGRLPLIDPSGISPKEAEGSSTEEAIT 180  
121 AITYERSAEGIYLGNGISPLGLHFRKALHFTGRLPLIDPSGISPKEAEGSSTEEAIT 180  
181 AGVIHGLAGEIDRYIDDLHAKGRSAVILTGGDANYLARIIRSGILIHDPDLVLLGLNRIL 240  
181 AGVIHGLAGEIDRYIDDLHAKGRSAVILTGGDANYLARIIRSGILIHDPDLVLLGLNRIL 240  
241 EYNY 244  
241 EYNY 244

RESULT 2  
AAU911153  
ID AAU911153 standard; Protein; 258 AA.  
XX  
AC AAU911153;  
XX  
DT 05-JUN-2002 (first entry)  
Rhodobacter capsulatus pantothenate kinase Coax.  
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX Rhodobacter capsulatus.  
OS  
XX WO200216601-A2.  
PN  
XX  
PD 28-FEB-2002.  
XX  
PF 24-AUG-2001; 2001WO-US26531.  
XX  
PR 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX  
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX  
PI Yocum RR, Patterson TA;  
XX  
DR WPI; 2002-269358/31.  
XX  
PT Identifying potential antibiotic or antimicrobial agent, comprises

PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein  
XX  
PS Claim 10; Page 71-72; 128pp; English.  
XX  
CC The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
XX described in the invention.  
SQ Sequence 258 AA;

Query Match 18.2%; Score 224; DB 23; Length 258;  
Best Local Similarity 29.4%; Pred. No. 1.1e-14;  
Matches 72; Conservative 37; Mismatches 86; Indels 50; Gaps 10;  
18 FVRNNSIESISFLPGKAGQALSHLVAP-----HRFDKAIYSSVGLPDEEAEIV 66  
40 FVWLNTLMQLKGLQGRISAEIISSTAPRVVFNLRVLCNRYFDCRPY-VVGKP----- 90  
67 RSCAAASLMMGTETPVPLRLQYDRRTLGAADRLAAVVGAAHSLYPNTELLVIDAGTAITYER 126  
91 -----GCELPVAPRVD-PGTTVGPDRLVNTVAGYDRH-GGDLIVVDFTGATTDFV 138  
127 VSAEGIYLGNGISPLGLHFRKALHFTGRLPLID---PSGISPKIAEYGSSTEEAITAGV 183  
139 VAPDGAYIGGVIAPIGVNLSLEALHMAAALPHVDVTKPGV-----IGTNTVACIQSGV 192  
184 IHGLAGEIDRYIDDLHAKGR-SAVILTGGDANYLARIIRSGI---LTHPDLVLLGLNR 238  
193 YWGIYGLVEGIVRQIRMERDRPMKVATGG----LASFLDLGFDLFDKVEDDLTMHGLRL 248  
239 ILEYN 243  
249 IFDYN 253

RESULT 3  
AAU911174  
ID AAU911174 standard; Protein; 257 AA.  
XX  
AC AAU911174;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Pantothenate kinase (Coax) #12.  
XX Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
KW Chlorobium tepidum.  
OS  
XX WO200216601-A2.  
PN  
XX 28-FEB-2002.  
PD  
XX 24-AUG-2001; 2001WO-US26531.  
PF  
XX 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX  
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX  
PI Yocum RR, Patterson TA;  
XX  
DR WPI; 2002-269358/31.  
XX  
PT N-PSDB; ABR54195.



OS Bacillus stearothermophilus.  
XX WO200216601-A2.  
PN 28-FEB-2002.  
XX  
PD  
XX  
PF 24-AUG-2001; 2001WO-US26531.  
XX  
PR 24-AUG-2000; 2000US-227860P.  
XX  
PR 20-MAR-2001; 2001US-0813453.  
XX  
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX  
PI Yocum RR, Patterson TA;  
XX  
DR WPI; 2002-269358/31.  
XX  
DR N-PSDB; ABK54193.

Identifying potential antibiotic or antimicrobial agent, comprises contacting composition comprising pantothenate kinase (Coax) protein with test compound and identifying inhibitor of the Coax protein -  
Claim 10; Page 101-102; 128pp; English.

The invention describes assays for identifying a (potential) antibiotic comprising contacting an assay composition comprising a pantothenate kinase (Coax) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulating agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (Coax) protein described in the invention.

Sequence 258 AA;

Query Match 17.1%; Score 210.5; DB 23; Length 258;  
Best Local Similarity 30.9%; Pred. No. 2.6e-13;  
Matches 68; Conservative 37; Mismatches 80; Indels 35; Gaps 10;  
QY 49 KATYSSVGLPDEEAEAIVRSCAAASLM-----MGTEPVPRLRLQYDR- 90  
Db 44 KALLNHVGLQFSDIRGIISVWPPIMFALERMCLKYFHKPLIVGPGIKTGLDIKYDNP 103  
QY 91 RTLGADRLA-AVUGAHSLYPNTLLVIDAGTAITYERSVAEGIYLGNNISPGHLRFRAL 149  
Db 104 REVGADRIVNAVAGIH-LY-GSPLIIVDFGTATTTCYINEHKQYMGGAIPGIMISTEAL 161  
150 HLFTGRRLPLID---PSGISPKIAEYGSSTEEAITAGVIHGLAGEIDRYIDDLHAKEG-RS 205  
162 FARAALKPRIETIARPDDI-----IGKNTVSAMQAGILYGVGVQVEGIVSRMKAISKIPP 215  
QY 206 AVILTGGDANYLARIIRSGIL--IHPDLVLLGLNRLILEYN 243  
Db 216 KVIATGGLAPLIAS--ESDIIDWVDPFLTLTGLKLLYEKN 253

RESULT 6  
AAU911154  
ID AAU911154 standard; Protein; 255 AA.  
XX  
AC AAU911154;  
XX  
DT 05-JUN-2002 (first entry)  
XX

Geobacter sulfurreducens pantothenate kinase Coax.  
Pantothenate kinase; Coax; antibiotic; antimicrobial;  
Pantothenate kinase modulator; coenzyme A; bactericidal compound.  
Geobacter sulfurreducens.

PN WO200216601-A2.  
XX 28-FEB-2002.  
PD  
XX  
PF 24-AUG-2001; 2001WO-US26531.  
XX  
PR 24-AUG-2000; 2000US-227860P.  
XX  
PR 20-MAR-2001; 2001US-0813453.  
XX  
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX  
PI Yocum RR, Patterson TA;  
XX  
DR WPI; 2002-269358/31.

Identifying potential antibiotic or antimicrobial agent, comprises contacting composition comprising pantothenate kinase (Coax) protein with test compound and identifying inhibitor of the Coax protein -  
Claim 10; Page 72-73; 128pp; English.

The invention describes assays for identifying a (potential) antibiotic comprising contacting an assay composition comprising a pantothenate kinase (Coax) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulating agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (Coax) protein described in the invention.

Sequence 255 AA;

Query Match 16.9%; Score 208; DB 23; Length 255;  
Best Local Similarity 34.1%; Pred. No. 4.7e-13;  
Matches 59; Conservative 34; Mismatches 64; Indels 16; Gaps 7;  
QY 77 GTETPVPRLRLQYDR-RTLGADRLAAVVGAAHSLYPNTLLVIDAGTAITYERSVAEGIYLG 135  
Db 91 GIKTGMP--IQYDNPREVGADRIVNAVAGYEKY-RTSLIIVDFGTATTFDYVNRKGEYCG 147  
QY 136 GNISPGHLRFRKALHLFTGRPLID---PSGISPKIAEYGSSTEEAITAGVIHGLAGEID 192  
Db 148 GAIAPGLVISTEALFORASKLPRVDIIRPSAIIAR-----NTVNSMQAGIYGYGVGLVD 201  
QY 193 RYIDDLHAKEGRSA--VILTGGDANYLARIIRSGILIHDPDLVLLGLNRLILEYN 243  
Db 202 EIVTRMKA-ESKDAPRVIAATGGLASLIAPESKTIIEAVEEYLTLEGLRILYERN 253

RESULT 7  
AAU911170  
ID AAU911170 standard; Protein; 262 AA.  
XX  
AC AAU911170;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Pantothenate kinase (Coax) #8.  
XX

Pantothenate kinase; Coax; antibiotic; antimicrobial;  
Pantothenate kinase modulator; coenzyme A; bactericidal compound.  
Bacillus anthracis.

WO200216601-A2.  
28-FEB-2002.  
24-AUG-2001; 2001WO-US26531.  
24-AUG-2000; 2000US-227860P.



PR 20-MAR-2001; 2001US-0813453.  
XX  
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX  
PI Yocum RR, Patterson TA;  
XX  
DR WPI; 2002-269358/31.  
DR N-PSDB; ABK54191.  
XX

Identifying potential antibiotic or antimicrobial agent, comprises contacting composition comprising pantothenate kinase (Coax) protein with test compound and identifying inhibitor of the Coax protein -  
Claim 8; Page 98-99; 128pp; English.  
The invention describes assays for identifying a (potential) antibiotic comprising contacting an assay composition comprising a pantothenate kinase (Coax) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulating agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (Coax) protein described in the invention.

Query Match 16.9%; Score 207.5; DB 23; Length 262;  
Best Local Similarity 30.8%; Pred. No. 5.5e-13;  
Matches 72; Conservative 44; Mismatches 73; Indels 45; Gaps 14;  
QY 25 ESISFLPGKAGQALSHLVAPHRFDKAIYSSVGLPDEEAIVRSC----AAASLMGTET 80  
Db 50 EGLSFEDVK-GIIVSSVPPIMF-----ALERMCEKFKIKPLVVGPGI 92  
QY 81 PVPLRLQYDR-RTLGDRLA-AVGAHSLYPNTELLVIDAGTAITYERSAEGIYLGNI 138  
Db 93 KTGLNIKENPREVGADRIVNAVAGIH-LY-GSPLIIVDFGTATTTCYINEEKHYMGVI 150  
QY 139 SPGLHLRFKALHFTGRPLID---PSGISPKIAEYSGSSTEEAITAGVIHGLAGEIDRYI 195  
Db 151 TPGMISAEALYSRAAKLPRIETKPSV-----VGKNTVSAMQSGILYGYVGQEGIV 204  
QY 196 DDL--HAKEGRSAVILTGGDANYLARI--RSGL--IHPDLVLLGLNRILEYN 243  
Db 205 KRMKEAKQ-EPKVIATGG----LAKLISEESNVIDVVDVDFLTCLKGLYMLYERN 253

RESULT 8  
AAU01243  
ID AAU01243 standard; Protein; 258 AA.  
XX  
AC AAU01243;  
XX  
DT 18-JUL-2001 (first entry)  
XX  
DE B. subtilis novel pantothenate kinase encoded by the gene coax.  
XX  
KW Pantothenate kinase; coax; pantothenate biosynthesis; vitamin B5;  
KW nutritional supplement; panto-compound; pantoate.  
XX  
OS Bacillus subtilis.  
XX  
PN WO200121772-A2.  
XX  
PD 29-MAR-2001.  
XX  
PF 21-SEP-2000; 2000WO-US25993.  
XX  
PR 21-SEP-1999; 99US-0400494.  
PR 07-JUN-2000; 2000US-0210072.  
PR 28-JUL-2000; 2000US-0221836.

PR 24-AUG-2000; 2000US-0227860.  
XX  
PA (OMNI-) OMNIGENE BIOPRODUCTS.  
XX  
PI Yocum RR, Patterson TA, Hermann T, Pero JG;  
XX  
DR WPI; 2001-218644/22.  
DR N-PSDB; AAS00984.  
XX

New recombinant microorganism which overexpress a Bacillus subtilis pantothenate biosynthetic enzyme, useful for the high yield production of panto-compounds such as pantothenate and pantoate -  
Example 14; Fig 23; 292pp; English.

The sequence represents a novel B. subtilis pantothenate kinase (encoded by gene coax), an enzyme of the pantothenate biosynthetic pathway. Pantothenate, also known as vitamin B5, is used as a nutritional supplement in mammals and humans. The invention concerns methods of producing recombinant microorganisms overexpressing at least one Bacillus subtilis pantothenate biosynthetic enzyme. The microorganisms and methods of producing them are useful for producing a panto-compound such as pantothenate or pantoate, which is a nutritional requirement for livestock and humans. The methods are also useful for the identification of pantothenate kinase modulators. Panto-compounds are produced at a significantly higher yield than prior art methods and can be produced independent of the need to feed precursors which decreases expense.

Query Match 16.8%; Score 207; DB 22; Length 258;  
Best Local Similarity 26.4%; Pred. No. 6.1e-13;  
Matches 71; Conservative 49; Mismatches 101; Indels 48; Gaps 10;  
QY 5 LIVDQGNACKVAFVRNNSIESISFLPGKAGQALSHLVAPHRFD-----KAIYSSVGL 57  
Db 3 LVIDVGNT-----NTVLGV-YHDGKLEYHWRIFTSRHKTEDEFGMILRSLEHSG 52  
QY 58 PDEEAIVRSCAAASLM-----MGTEPVPVLRLOQYDR-RTLGDRLA 99  
Db 53 MFEQIDGIIISVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKIDNPKEVGADRIV 112  
QY 100 AVVGAHSLYPNTELLVIDAGTAITYERSAEGIYLGNI SPGLHLRFKALHFTGRPLI 159  
Db 113 NAVAAIHLYGN-PLIVVDFGTATTTCYIDENKQYMGGAIPGITISTEALYSRAAKLPRI 171  
QY 160 D---PSGISPKIAEYSGSSTEEAITAGVIHGLAGEIDRYIDDL--HAKEGRSAVILTGG 214  
Db 172 EITRPDNI-----IGKNTVSAMQSGILFGYVGQEGIVKRMKQAKQDLK-VIATGG 224  
QY 215 NYLARIIRSGILIHPLDLVLLGLNRILEYN 243  
Db 225 PLIANESDCIDIVDFLTCLKGLELIYERN 253

RESULT 9  
AAU91149  
ID AAU91149 standard; Protein; 258 AA.  
XX  
AC AAU91149;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Bacillus subtilis pantothenate kinase Coax.  
XX  
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX  
OS Bacillus subtilis.  
XX  
PN WO200216501-A2.  
XX  
PD 28-FEB-2002.

```
XX PF 24-AUG-2001; 2001WO-US26531.
XX PF
XX PR 24-AUG-2000; 2000US-227860P.
XX PR 20-MAR-2001; 2001US-0813453.
XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX PI Yocum RR, Patterson TA;
XX PI WPI; 2002-269358/31.
XX DR N-PSDB; ABK54168.
XX DR
XX PT Identifying potential antibiotic or antimicrobial agent, comprises
XX PT contacting composition comprising pantothenate kinase (Coax) protein
XX PT with test compound and identifying inhibitor of the Coax protein
XX PS Claim 10; Page 67-68; 128pp; English.
XX CC The invention describes assays for identifying a (potential) antibiotic
XX CC comprising contacting an assay composition comprising a pantothenate
XX CC kinase (Coax) protein with a test compound, and determining the ability
XX CC of the test compound to inhibit the activity of the Coax protein, an
XX CC essential enzyme for the production of coenzyme A. Coax protein is a
XX CC valuable target for identifying bactericidal compounds. Coax modulating
XX CC agents can be used in an infectious animal model to determine the
XX CC efficacy, toxicity, or side effects of treatment with such an agent. This
XX CC is the amino acid sequence of a pantothenate kinase (Coax) protein
XX CC described in the invention.
XX SQ Sequence 258 AA;
Query Match 16.8%; Score 207; DB 23; Length 258;
Best Local Similarity 26.4%; Pred. No. 6.1e-13;
Matches 71; Conservative 49; Mismatches 101; Indels 48; Gaps 10;
QY 5 LIVDQNSACKVAFVRNNSIESISFLPGKAGQALSHLVAPHRFD-----KAIYSSVGL 57
Db 3 LVIDVGN-----NTVLGV-YHDGKLEYHWRIETSRHKTEDEFGMILRSLEDSGL 52
QY 58 PDEEAIVRSCAAASLM-----MGTETPVPLRLQYDR-RTLQADRLA 99
Db 53 MFEQIDGIIISVVPPINFALERMCTKYFHIEPQIVGPMKTLGNIKYDNPKEVGADRIV 112
QY 100 AVVGAHSLYPNTELLVIDAGTAITYERSVSAEGIYLGNSPGLHLRFRKALHFTGRLPLI 159
Db 113 NAVAIAHLGN-PLIVVDFGTATTCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRI 171
QY 160 D---PSGISPKIAEYSGSTEEAITAGVIHGLAGEIDRYIDDL--HAKEGRSAVILTGDDA 214
172 EITRPDNI-----IGKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKODLK-VIATGGLA 224
QY 215 NYLARIIRSGILIHPLDLVLLGLNRILEYN 243
Db 225 PLIANESDCIDIVDPFTLKGLELIYERN 253
RESULT 10
AAU91151
ID AAU91151 standard; Protein; 265 AA.
XX AC AAU91151;
XX XX
XX DT 05-JUN-2002 (first entry)
XX DE Streptomyces coelicolor pantothenate kinase Coax.
XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX OS Streptomyces coelicolor.
XX XX
XX PN WO200216601-A2.
```

```
XX PD 28-FEB-2002.
XX PF
XX PF 24-AUG-2001; 2001WO-US26531.
XX PR 24-AUG-2000; 2000US-227860P.
XX PR 20-MAR-2001; 2001US-0813453.
XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX PI Yocum RR, Patterson TA;
XX PI WPI; 2002-269358/31.
XX DR
XX PT Identifying potential antibiotic or antimicrobial agent, comprises
XX PT contacting composition comprising pantothenate kinase (Coax) protein
XX PT with test compound and identifying inhibitor of the Coax protein
XX PS Claim 10; Page 69-70; 128pp; English.
XX CC The invention describes assays for identifying a (potential) antibiotic
XX CC comprising contacting an assay composition comprising a pantothenate
XX CC kinase (Coax) protein with a test compound, and determining the ability
XX CC of the test compound to inhibit the activity of the Coax protein, an
XX CC essential enzyme for the production of coenzyme A. Coax protein is a
XX CC valuable target for identifying bactericidal compounds. Coax modulating
XX CC agents can be used in an infectious animal model to determine the
XX CC efficacy, toxicity, or side effects of treatment with such an agent. This
XX CC is the amino acid sequence of a pantothenate kinase (Coax) protein
XX CC described in the invention.
XX SQ Sequence 265 AA;
Query Match 16.7%; Score 205.5; DB 23; Length 265;
Best Local Similarity 33.5%; Pred. No. 9e-13;
Matches 59; Conservative 36; Mismatches 64; Indels 17; Gaps 7;
QY 77 GTETPVPLRLQYDRRTLGADRLAAVVGHAHSLYPNTELLVIDAGTAITYERSVSAEGIYLG 136
Db 96 GVKTGVPILTDHPKE-VGADRIINAAVAVELY-GGPAIVDFGTATTFDAVSARGEYIGG 153
QY 137 NISPLHLRFRKALHFTGRLPLIDPSGISPKIAEYSGSTEEAITAGVIHGLAGEID---- 192
Db 154 VIAPGIEISVEALGVKGAQLRKIE--VARPRSVIGKNTVERMQSGIYVGFAGQVDGVN 210
QY 193 RYIDDLHAKEGRSAVILTGDDANLARI--RSGIL--IHPDLVLLGLNRILEYNV 244
Db 211 RMARELADDDDDVTVIATGG-----LAPMVLGESSVIDEHEPWLTLMLRLVYERNV 262
RESULT 11
AAU91175
ID AAU91175 standard; Protein; 256 AA.
XX AC AAU91175;
XX XX
XX DT 05-JUN-2002 (first entry)
XX DE Pantothenate kinase (Coax) #13.
XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX OS Clostridium difficile.
XX XX
XX PN WO200216601-A2.
XX PD 28-FEB-2002.
XX PF 24-AUG-2001; 2001WO-US26531.
XX PR 24-AUG-2000; 2000US-227860P.
XX PR 20-MAR-2001; 2001US-0813453.
```

XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX PA Yocum RR, Patterson TA;  
XX PI WPI; 2002-269358/31.  
XX DR N-PSDB; ABK54196.

XX Identifying potential antibiotic or antimicrobial agent, comprises  
XX contacting composition comprising pantothenate kinase (Coax) protein  
XX with test compound and identifying inhibitor of the Coax protein -  
PS Claim 6; Page 105; 128pp; English.  
XX The invention describes assays for identifying a (potential) antibiotic  
XX comprising contacting an assay composition comprising a pantothenate  
XX kinase (Coax) protein with a test compound, and determining the ability  
XX of the test compound to inhibit the activity of the Coax protein, an  
XX essential enzyme for the production of coenzyme A. Coax protein is a  
XX valuable target for identifying bactericidal compounds. Coax modulating  
XX agents can be used in an infectious animal model to determine the  
XX efficacy, toxicity, or side effects of treatment with such an agent. This  
XX is the amino acid sequence of a pantothenate kinase (Coax) protein  
XX described in the invention.

SQ Sequence 256 AA;

Query Match 16.4%; Score 201.5; DB 23; Length 256;  
Best Local Similarity 27.0%; Pred. No. 2.2e-12;  
Matches 75; Conservative 44; Mismatches 94; Indels 65; Gaps 11;  
QY 5 LIVDQNSACKVAFVRNNSI-----ESISFLPGKAGQALSHLVAPHRFDKAIYSS 54  
Db 3 LVFDVGNTNMVLGIYKDKLVNYWRIKTDREKTS---DEYGILISNL-----FD---YDN 51  
QY 55 VGLPDEEAEL-----VRSAAASLMMGTETPVPLRLQYDR-RTLGDRL 98  
Db 52 VNISDIDDVLISSVVPNMVHSLNFCEIKYCKKQPLIVGPGIKTGLNICYNDPKQVGADRI 111  
QY 99 AAVGASHLPNTLELLVIDAGTAITYYERVSAGIYLGGNISPGHLRFRKALHFTGRPL 158  
Db 112 VNAVAGIEKYGAPSLV-DFGTATTFCAISEKGEYLGGTIAPGKISSEALFQSASKLPR 170  
QY 159 IDPSGISPKIAEYG---SSTEEAITAGVIHGLAGEIDRYI-----DDLHAKEGRS 205  
Db 171 VE-----LAKPGMTICKSTVSAMQSGIYGYVGLVDKIIISIMKKELNCDV----- 216

QY 206 AVILTGGDANYLARIIRSGILIHDPDLVLLGLNRILEYN 243  
Db 217 KVIATGGLAKLIASETKSIDYVDGFLTLEGLRIIYEKN 254

RESULT 12  
AAU91173  
ID AAU91173 standard; Protein; 260 AA.  
XX AC AAU91173;  
XX DT 05-JUN-2002 (first entry)  
XX DE Pantothenate kinase (Coax) #11.

XX Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW Pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX OS Caulobacter crescentus.  
XX PN WO200216601-A2.  
XX PD 28-FEB-2002.  
XX PF 24-AUG-2001; 2001WO-US26531.

PR 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX Yocum RR, Patterson TA;  
XX WPI; 2002-269358/31.  
XX DR N-PSDB; ABK54194.

XX Identifying potential antibiotic or antimicrobial agent, comprises  
XX contacting composition comprising pantothenate kinase (Coax) protein  
XX with test compound and identifying inhibitor of the Coax protein -  
PS Claim 10; Page 102-103; 128pp; English.

XX The invention describes assays for identifying a (potential) antibiotic  
XX comprising contacting an assay composition comprising a pantothenate  
XX kinase (Coax) protein with a test compound, and determining the ability  
XX of the test compound to inhibit the activity of the Coax protein, an  
XX essential enzyme for the production of coenzyme A. Coax protein is a  
XX valuable target for identifying bactericidal compounds. Coax modulating  
XX agents can be used in an infectious animal model to determine the  
XX efficacy, toxicity, or side effects of treatment with such an agent. This  
XX is the amino acid sequence of a pantothenate kinase (Coax) protein  
XX described in the invention.

SQ Sequence 260 AA;

Query Match 16.1%; Score 198; DB 23; Length 260;  
Best Local Similarity 33.7%; Pred. No. 5.3e-12;  
Matches 59; Conservative 34; Mismatches 72; Indels 10; Gaps 6;  
QY 74 LMMGTETPVPLRLQYDRRT-LGADRLAAVVGASHLYPNTELLVIDAGTAITYYERVSAGI 132  
Db 86 LVIGENAKLGIDVRIEKPESEAGADRLVNAIGAAMVYPG-PLVVIDSGTATTFDVAADGA 144  
QY 133 YLGGNISPGHLRFRKALHFTGRPLI---DPSGISPKIAEYGSSTEEAITAGVIHGLAG 189  
Db 145 FEGGIAPGINLSMQALHEAAKLPRIAIQRPAG--NRIV--GTDVVSAMQSGVFWGYIS 200  
QY 190 EIDRYIDDLHAKEGR-SAVILTGGDANYLARIIRSGILIHDPDLVLLGLNRILEYN 243  
Db 201 LIEGLVARIKAERGEPTVIATGCVASLFEGATDSIDHFDSDITIRGLLEIYRRN 255

RESULT 13  
AAU91171  
ID AAU91171 standard; Protein; 254 AA.  
XX AC AAU91171;  
XX DT 05-JUN-2002 (first entry)  
XX DE Pantothenate kinase (Coax) #9.  
XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW Pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX OS Bacillus halodurans.  
XX PN WO200216601-A2.  
XX PD 28-FEB-2002.  
XX PF 24-AUG-2001; 2001WO-US26531.  
XX PR 24-AUG-2000; 2000US-227860P.  
XX PD 20-MAR-2001; 2001US-0813453.  
XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX Yocum RR, Patterson TA;

XX WPI; 2002-269358/31.  
DR N-PSDB; ABK54192.  
XX  
PT Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein  
XX  
PS Claim 10; Page 100; 128pp; English.  
XX  
CC The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.  
XX  
SQ Sequence 254 AA;

Query Match 16.1%; Score 197.5; DB 23; Length 254;  
Best Local Similarity 27.9%; Pred. No. 5.7e-12;  
Matches 60; Conservative 44; Mismatches 88; Indels 23; Gaps 6;  
QY 49 KATYSSVGLPDEEAIAIVSCAAASLMGTE-----TPV-----PLRLQYDR- 90  
Db 44 RSLFDHAGLOFQDIDGIVISSVPPMMFSLQMKCKYFHTPMLIGPGIKTGLNIKYDNP 103  
QY 91 RTLGDRLAAVVGASLNPTELLVIDAGTAITYERSAEGIYLGGNISPLHLRFRKALH 150  
Db 104 KEVGADRIVNAVAIELY-GYPAIVVDFGTATTYCLINEKKQYAGGVIAPIGIMISTEALY 162  
QY 151 LFTGRLPLIDPSGISPKIAEYGSSTEEAITAGVIHGLAGEIDRYIDDLHAK-EGRSVAIL 209  
Db 163 HRASKLPRIE---IAKPKQVVGTNTIDSMQSGIFGYVSQVDGWWKRMKAQAESEPKVIA 219  
QY 210 TGGDANYLARIIRSGILIHDPDLVLLGLNRILEYV 244  
Db 220 TGGLAKLIGTESETIDVIDSFLTLKGLQLIYKKNV 254

RESULT 14  
AAU911164  
ID AAU911164 standard; Protein; 248 AA.  
XX  
AC AAU911164;  
05-JUN-2002 (first entry)  
DE Pantothenate kinase (Coax) #2.  
XX  
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX  
OS Pseudomonas aeruginosa.  
XX  
PN WO200216601-A2.  
XX  
PD 28-FEB-2002.  
XX  
PF 24-AUG-2001; 2001WO-US26531.  
XX  
PR 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX  
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX  
PI Yocum RR, Patterson TA;  
XX  
DR WPI; 2002-269358/31.

XX Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein  
XX  
PS Claim 6; Page 83-84; 128pp; English.  
XX  
CC The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.  
XX  
SQ Sequence 248 AA;

Query Match 15.9%; Score 195.5; DB 23; Length 248;  
Best Local Similarity 30.4%; Pred. No. 8.9e-12;  
Matches 80; Conservative 33; Mismatches 95; Indels 55; Gaps 12;  
QY 5 LIVDQNSACKVAFVRNNSIESISFLPGKAGQALSHLVAPHRFDKAIYSSVGLPDEEA 64  
Db 3 LEIDCGNSLIKWRVI-EGAARSAV---GGLAESDDALV-----EQLTSQOALP----- 46  
QY 65 IVRSCAAASLMMGTE-----PVP-----LRLQY-DRRTLGADRLAA 100  
Db 47 -VRACRLVSRSEQETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYORLGLDRWLA 105  
QY 101 VVGAHSLYPNTELLVIDAGTAITYERSAEGIYLGGNISPLHLRFRKALHFTGRPLID 160  
Db 106 LVAAHHL-AKKACLVIDLGTAVTSDLVADGVHLGGYICPGMTLMRSQRLTHRRIRYDD 164  
QY 161 P-----SGISPKIAEYGSSTEEAITAGVIHGLAGEI-DRYIDDLHAKEGRSVAILTGGD 213  
Db 165 AEARRALASLP-----GQATAEAVERGCLMLRGFVREQYAMACELLPDCEIFLTGGD 219  
QY 214 ANYLARIIRSGILIHDPDLVLLGL 236  
Db 220 AE-LVRDELAGARIMPDVLFVGL 241

RESULT 15  
AAU911176  
ID AAU911176 standard; Protein; 219 AA.  
XX  
AC AAU911176;  
05-JUN-2002 (first entry)  
DE Pantothenate kinase (Coax) #14.  
XX  
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX  
OS Dehalococcoides ethenogenes.  
XX  
PN WO200216601-A2.  
XX  
PD 28-FEB-2002.  
XX  
PF 24-AUG-2001; 2001WO-US26531.  
XX  
PR 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX  
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX  
PI Yocum RR, Patterson TA;  
XX



Search completed: June 24, 2003, 21:46:29  
Job time : 26.051 secs.

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:36:04 ; Search time 8.8997 Seconds  
(without alignments)  
2635.685 Million cell updates/sec

Title: US-09-813-453A-41  
Perfect score: 1229  
Sequence: 1 MSFNLIVDQGNACKVAFVR.....ILIHFDLVLLGLNRILEYNV 244

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	216	17.6	259	2 AF1102	conserved hypothet
2	214	17.4	259	2 AF1464	conserved hypothet
3	208	16.9	273	2 E97293	probable transcrip
4	205.5	16.7	265	2 T36391	hypothetical prote
5	198	16.1	261	2 B87489	transcription acti
6	197.5	16.1	254	2 F83660	hypothetical prote
7	195.5	15.9	248	2 H83111	hypothetical prote
8	183	14.9	276	2 A12292	conserved hypothet
9	181.5	14.8	242	2 A82637	conserved hypothet
10	175.5	14.3	233	2 S66100	conserved hypothet
11	172.5	14.0	273	2 D71326	conserved hypothet
12	163	13.3	267	2 I40327	baf protein - Bord
13	162	13.2	246	2 D72320	conserved hypothet
14	161	13.1	592	2 B81009	BirA protein/Bvg a
15	157.5	12.8	274	2 H86937	conserved hypothet
16	156	12.7	592	2 H82031	probable biotin-la
17	153	12.4	272	2 A70955	hypothetical prote
18	152	12.4	257	2 S75559	hypothetical prote
19	135	11.0	262	2 F70165	conserved hypothet
20	134	10.9	223	2 F71887	hypothetical prote
21	132.5	10.8	229	2 E70465	hypothetical prote
22	131	10.7	262	2 E75516	conserved hypothet
23	129.5	10.5	223	2 F64627	hypothetical prote
24	126.5	10.3	209	2 H81382	hypothetical prote
25	113	9.2	450	2 C95348	HemN coproporphyr
26	105.5	8.6	456	2 JC4089	3alpha,7alpha,12al
27	96.5	7.9	385	2 G87340	aminotransferase,
28	96	7.8	350	2 A75600	conserved hypothet
29	95.5	7.8	240	2 G91045	hypothetical prote

30	95.5	7.8	240	2 C85890	hypothetical prote
31	95.5	7.8	240	2 H65022	hypothetical prote
32	95.5	7.8	468	1 S19723	dihydrolipoamide d
33	95	7.7	332	2 G71297	probable fructose-
34	94.5	7.7	534	2 A98326	ketoglutarate semi
35	94.5	7.7	534	2 AE2957	dehydrogenase Atu3
36	94	7.6	298	2 F84224	hypothetical prote
37	93	7.6	323	2 H86765	L-lactate dehydrog
38	93	7.6	559	2 C87307	hypothetical prote
39	91.5	7.4	342	2 AB3169	conserved hypothet
40	90.5	7.4	758	2 T02925	protoporphyrin IX
41	90	7.3	1168	2 E70726	probable polyketid
42	89.5	7.3	735	2 T08140	1-deoxy-D-xylulose
43	89	7.2	382	2 AG0567	glycerate kinase {
44	89	7.2	467	2 AD3270	dihydrolipoamide d
45	88.5	7.2	409	2 H95911	probable ferredoxi

ALIGNMENTS

RESULT 1

AF1102

conserved hypothetical protein lmo0221 [imported] - Listeria monocytogenes (strain E  
C:Species: Listeria monocytogenes  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C:Accession: AF1102  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloe  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.  
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehl  
A:Title: Comparative genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AF1102

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-259 <GLA>

A:Cross-references: GB:NC\_003210; PIDN:CAD00748.1; PID:g16409586; GSPDB:GN00177

A:Experimental source: strain EGD-e

C:Genetics:

A:Gene: lmo0221

C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 17.6%; Score 216; DB 2; Length 259;  
Best Local Similarity 25.4%; Pred. No. 1.7e-10;  
Matches 69; Conservative 43; Mismatches 106; Indels 54; Gaps 9;

Qy 5 LIVDQNSACKVAFVRNNSIESISFLPGKAGQALSH---LVAPHRFDKAI-----YSS 54

Db 3 LVIDVGNCTCTGVYEKQKL-----LKHWRMTTDRHRTSDELGMTVLNFFSY 49

Qy 55 VGLPDERAEAIVRSCAAASLMGMTET-----PVPLRLQYDR-RTLGD 96

Db 50 ANLTSPDIQGIISVVPPIMHAMETMCVRYFNIRPLIVPGIKTGLNLKVDNPREIGSD 109

Qy 97 RLAAVVGASLYPNTPELLVIDAGTAITYERSVSAEGIYLGNNISPGHLRFKALHLFTGRL 156

Db 110 RIVNAVAASEEY-GTPVIVDFGTATTCYIDESGVYQGGAIAPIGIMISTEALYNRAKL 168

Qy 157 PLIDPSGISPKIAE---YGSSTEEAITAGVIHGLAGEIDRYIDDLHAKGRSAVIL-TG 211

Db 169 PRVD-----IAESSQIIGKSTVSSMQAGIFYGVFGQCEGIIAEMKKOSNASPVVAVTG 221

Qy 212 GDANYLARIIRSGILIHFDLVLLGLNRILEYN 243

Db 222 GLARMIKSSAVDILDPLFTLTKGLELLYRRN 253

RESULT 2

AF1464

conserved hypothetical protein lin0253 [imported] - Listeria innocua (strain Clip112





```

Db      87 LVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAMVYPG-PLVVIDSGTATTFDIVAADGA 145
Qx     133 YLGGNISPGHLRFAKALHFTGRLPLI--DPSGISPKIAEYGSSTEEAITAGVIHGLAG 189
Db     146 FEGGIIAPGINLSMQALHEAAAKLPRIAQRPAQ--NRIV--GTDVTSAMQSGVFWGYIS 201
QY     190 EIDRYIDDLHAKEGR-SAVILTGGDANYLARIIRSGILIHDPDLVLLGLNRILEYN 243
Db     202 LIEGLVARIKAERGEPMTVIATGGVASLFEGATSDIDHFDSDLTIRGLEIYRN 256

RESULT 6
F83660
hypothetical protein BH0086 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: F83660
r;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: F83660
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-254 <STO>
A;Cross-references: GB:AP001507; GB:BA000004; NID:g10172612; PIDN:BA03805.1; GSPDB:GN00
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH0086
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match      16.1%; Score 197.5; DB 2; Length 254;
Best Local Similarity 27.9%; Pred. No. 5.7e-09;
Matches 60; Conservative 44; Mismatches 88; Indels 23; Gaps 6;

QY     49 KAIYSSVGLPDEAEAIVRSCAAASLMMGTE-----TPV-----PLRLQYDR- 90
Db     44 RSLFDHAGLQFDQDIDGIVISSVPPMFESLEQMCKKYFHVTPMIIGPGIKTGLNIKYDNP 103

QY     91 RTLGADRLAAVVGAAHSLYPNTELLVIDAGTAITYERVSAEGIYLGNNISPLHLRFAKALH 150
Db     104 KEVGADRIVNAAVAIELY-GYPAIVVDFGTATTYCLINEKKQYAGGVIAAPGIMISTEALY 162

QY     151 LFTGRLPLIDPSGISPKIAEYGSSTEEAITAGVIHGLAGEIDRYIDDLHAK-EGRSAVIL 209
Db     163 HRASKLPRIE---IAKPKQVGTNTIDSMQSGIFYGYVSQVDGVVKKRKAQAESEPKVIA 219

QY     210 TGGDANYLARIIRSGILIHDPDLVLLGLNRILEYNNV 244
Db     220 TGGLAKLIGTESETIDVIDSFLTGLQLIYKKNV 254

RESULT 7
H83111
hypothetical protein PA4279 [imported] - Pseudomonas aeruginosa (strain PA01)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: H83111
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: H83111
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-248 <STO>
A;Cross-references: GB:AE004843; GB:AE004091; NID:g9950489; PIDN:AAG07667.1; GSPDB:GN001
A;Experimental source: strain PA01
C;Genetics:
A;Gene: PA4279
```

```

Query Match      15.9%; Score 195.5; DB 2; Length 248;
Best Local Similarity 30.4%; Pred. No. 8.1e-09;
Matches 80; Conservative 33; Mismatches 95; Indels 55; Gaps 12;

QY     5 LIVDQGSACKVAFVRNNSSTESISFLPGKAGQALSHLVAPHRFDKAIYSSVGLPDEAEA 64
Db     3 LELDCGNSLIKRWI-EGAARSAVA---GGLAESDDALV-----EQLTSQALP----- 46

QY     65 IVRSCAAASLMMGTET-----PVP-----LRLOY-DRRTLGADRLAA 100
Db     47 -VRACRLVSVRSEQETSQVLVARLEQLFPVSALVASSGKQLAGVNGYLDYQRLGLDRWLA 105

QY     101 VVGAHSLYPNTELLVIDAGTAITYERVSAEGIYLGNNISPLHLRFAKALHFTGRLPLID 160
Db     106 LVAAHHL-AKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHRRIRYDD 164

QY     161 P-----SGISPKIAEYGSSTEEAITAGVIHGLAGEI-DRYIDDLHAKEGRSAVILTGGD 213
Db     165 AEARRALASLQP-----GQATAEAVERGCLLMRLRGFVREQYAMACELLGPDCEIFLTGGD 219

QY     214 ANYLARIIRSGILIHDPDLVLLGL 236
Db     220 AE-LVRDELAGARIMPDVFEVGL 241

RESULT 8
AI2292
hypothetical protein alr3896 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp.
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C;Accession: AI2292
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iri
Nakazaki, N.; Shimpō, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Taba
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacteriu
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AI2292
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-276 <KUR>
A;Cross-references: GB:BA000019; PIDN:BA075595.1; PID:g17133030; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr3896

Query Match      14.9%; Score 183; DB 2; Length 276;
Best Local Similarity 31.3%; Pred. No. 1e-07;
Matches 52; Conservative 33; Mismatches 71; Indels 10; Gaps 4;

QY     82 VPLRLQYDRRTLGAADRLAAVVGAAHSLYPNTELLVIDAGTAITYERVSAEGIYLGNNISPG 141
Db     110 IPLNNIYP--TLGIDRALALWAGMSW-GFPVLVIDAGTALTFTAADGGKNLVGGAILPG 166

QY     142 LHLRFKALHFTGRLPLIDPSGISPKIAEYGSSTEEAITAGVIHGLAGEIDRYIDDLHAK 201
Db     167 VGLQFASLGGQQTQLPQVEMEAIKSLPPRFALNTTEAIQSGVIYTLIAGMRDFTTEWLSL 226

QY     202 EGRSAVILTGGD---ANYLARI---IRSGILIHDPDLVLLGLNRIL 240
Db     227 FPDGKVAIKGGDRILLNLYQALYPDLAARLIVEPNLIFWGMQTIV 272

RESULT 9
A82637
conserved hypothetical protein XF1795 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: A82637
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide S
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
```

Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,

114 IVIDAGTATTVBRVSAEFTYI GGNTSPGLHLREKALHL ETCRLPLIDPSGTS PKIAEYGS 17

Db 131 VVDCGTALTFTAVDGTGLIQGVATAPGLRTAVQSLHTGTQAQLPLV-PLALPDSV--LGK 187

QY 174 STEEAITAGVIHGLAGEIDRYIDDLHAKEG-RSAVILTTGGDANVLIARIIRSGILIHLP--- 229

Db 188 DTHAVQAGVVRGTLFVIRAMIAQCQKELGCRCAAVITGG----LSRLFSSEVDFPPIDA 243

QY 230 DLVLLGLNRI 239

Db 244 QLTLSGLAHI 253

RESULT 12

I40327

baf protein - Bordetella pertussis

C;Species: Bordetella pertussis

C;Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 08-Oct-1999

C;Accession: I40327; S70669

R;DeShazer, D.; Wood, G.E.; Friedman, R.L.

Bacteriol. 177, 3801-3807, 1995

Title: Identification of a Bordetella pertussis regulatory factor required for transcr

A;Reference number: I40327; MUID:95325323; PMID:7601846

A;Accession: I40327

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-267 <RES>

A;Cross-references: EMBL:U12020; NID:g687228; PIDN:AAA75361.1; PID:g687229

R;Allen, A.; Maskell, D.

Mol. Microbiol. 19, 37-52, 1996

A;Title: The identification, cloning and mutagenesis of a genetic locus required for lip

A;Reference number: S70669; MUID:96419162; PMID:8821935

A;Accession: S70669

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 239-267 <ALL>

A;Cross-references: EMBL:X90711; NID:g992967; PIDN:CAA62242.1; PID:g992968

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

C;Genetics:

A;Gene: baf

Query Match 13.3%; Score 163; DB 2; Length 267;

Best Local Similarity 30.8%; Pred. No. 4.3e-06;

Matches 66; Conservative 28; Mismatches 78; Indels 42; Gaps 10;

QY 5 LIVDQNSACKVAFVRNNSIESISFLPGKAGQALSHLVAPHRFDKAIYSSVG-----LPD 59

Db 3 ILIDSGNSRLKVGW-----FDP-DAPQA-AREPAPVAFDNLDDLGRWLATLPR 50

QY 60 EEAEALVRSCA-----AASLMMG-----TETPVPLRLQYDRRT---LGADRLAAV 101

Db 51 RPQRALGVNVAGLARGEIAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDLQADRWACM 110

QY 102 VGAHSLYPNTE--LLVIDAGTAITYERSVAEGIYLGGNISPGHLRFKALHLFTGRLPLI 159

Db 111 VGVLARQPSVHPPLLVASFGTATLDTIGPDNVFPGGLILPGPAMRGALAYGTAHLPLA 170

QY 160 DPSGISPKIAEYGSSTEEAITAGVIHGLAGEIDR 193

Db 171 D--GL--VADYPIDTHQAIASGIAAAQAQGAIVR 199

RESULT 13

D72320

conserved hypothetical protein - Thermotoga maritima (strain MSB8)

C;Species: Thermotoga maritima

C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000

C;Accession: D72320

R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey

Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;

C.M.

Nature 399, 323-329, 1999

A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq

A;Reference number: A72200; MUID:99287316; PMID:10360571

A;Accession: D72320

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-246 <ARN>

A;Cross-references: GB:AE001754; GB:AE000512; NID:g4981417; PIDN:AAD35964.1; PID:g4;

A;Experimental source: strain MSB8

C;Genetics:

A;Gene: TM0883

C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 13.2%; Score 162; DB 2; Length 246;

Best Local Similarity 33.3%; Pred. No. 4.7e-06;

Matches 48; Conservative 23; Mismatches 67; Indels 6; Gaps 4;

QY 93 LGADRLAAVVGAHSLYPNTELLVIDAGTAITYERSVAEGIYLGGNISPGHLRFKALHLF 152

Db 102 VGADRVANVAVFVKEYGKNG-IIIDMGATTATVDLV-VNGSYEGGAILPGFFMMVHSLFRG 159

QY 153 TGRPLIDPSGISPKIAEYGSSTEEAITAGVIHGLAGEIDRYIDDLHAKEGRSVILTTGG 212

Db 160 TAKLPLVE---VKPADFVVGKDFTEENIRLGVVNGSVYALEGIIGRIKEYGDLPPVLTGG 216

QY 213 DANYLARIIRSGILIHDPDLVLLGL 236

Db 217 QSKIVKDMIKHEI-FDEDLTIKGV 239

RESULT 14

B81009

Bira protein/Bvg accessory factor NMB2075 [imported] - Neisseria meningitidis (strai

C;Species: Neisseria meningitidis

C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001

C;Accession: B81009

R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,

Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, E

ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.

Science 287, 1809-1815, 2000

A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.

A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.

A;Reference number: A81000; MUID:20175755; PMID:10710307

A;Accession: B81009

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-592 <TET>

A;Cross-references: GB:AE002557; GB:AE002098; NID:g7227332; PIDN:AAF42394.1; PID:g7;

A;Experimental source: serogroup B, strain MC58

C;Genetics:

A;Gene: NMB2075

Query Match 13.1%; Score 161; DB 2; Length 592;

Best Local Similarity 25.0%; Pred. No. 1.8e-05;

Matches 65; Conservative 43; Mismatches 102; Indels 50; Gaps 9;

QY 5 LIVDQNSACKVAFVRNNSIESISFLPGKAGQALSHLVAPHRFDKAIYSSVGLPDEEAE 64

Db 341 LLLDGGNSRLKNAWVENGTFATVGSAPYR---DLSPLGAEWA-----EKADG 384

QY 65 IVR--SCAAA-----SLMMGTETPVPLRLQYDR-RTLGAADRLAAVVGAH 105

Db 385 NVRIVGCAVCGEFKKAQVQEQARKIEWLPSSAQLGIRNHYRHPHEHSGDRWFNALGSR 444

QY 106 SLYPNTELLVIDAGTAITYERSVAEGIYLGGNISPGHLRFKALHLFTGRLPLIDPSGIS 165

Db 443 RFSRNA-CVVVSCGTAVTDALTDGHHYLGTTIMPGFHLMKESLAVRTANL-----NRHA 498

QY 165 PKIAEYGSSTEEAITAGVIHGLAGEIDRYIDDLHAKEGRSA---VILTTGGDANYLARIIR 222

Db 499 GKRYFPPTTTGNVAVSGMMDAVCGSVMMHGRLEKTKGAGKPVVDVIITGGAAKVAEALP 558

QY 223 SGIL-----IHPDLVLLGL 236

Db 559 PAFLAENTVRVADNLVIYGL 578

## H86937

Query Match	12.8%;	Score 157.5;	DB 2;	Length 274;
Best Local Similarity	31.5%;	Pred. No. 1.3e-05;		
Matches	58;	Conservative 29;	Mismatches 66;	Indels 31; Gaps 9;

  

QY	77	GTETPVPLRLOYDRRTL	GADRLAAVVG	AHSLYPNT	ELLVIDAGTA	ITYERVS	AEGIVLGG	136		
		:     :	: :      :	:	:    :    :	:    :	:			
Dd	93	GVRTGIPL-LVDNPK	EVGADRIVNCL	AAFKF-GQAA	IVVDFGSS	ICVDVVS	AKGEFLGG	150		
		:     :	: :      :	:	:    :    :	:    :	:			
QY	137	NISPGHL-----	RFKALHFT	GTGRPL	IDPDSG	ISPKIAEY	GSSTEEA	ITAGVIHGLAG	189	
		:     :	: :      :	:	:    :    :	:    :	:			
Dd	151	AIAPGVQVSSDAA	AARSALR----	RVELAR	PRSV-----	VGKNT	VECMQAGV	VVEGFAG	200	
		:     :	: :      :	:	:    :    :	:    :	:			
QY	190	EID----	RYIDDLHAK	EG-----	RSAVIL	TGGDAN	YLARIIR	SGILIHDP--	LVLLGLNRI	239
		:     :	: :      :	:	:    :    :	:    :	:	:    :	:    :	
Dd	201	LVDGLVGRMRQD	VEEFSGDL	GNRVAV	VATGHT	APLLPE	LHT--	VDHYDR	HLTLHGLRLV	258
		:     :	: :      :	:	:    :    :	:    :	:	:    :	:    :	
QY	240	LEYN	243							
		:     :	: :      :	:	:    :    :	:    :	:	:    :	:    :	
Dd	259	FERN	262							
		:     :	: :      :	:	:    :    :	:    :	:	:    :	:    :	

Search completed: June 24, 2003, 22:03:22  
Job time : 9.8997 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:33:11 ; Search time 4.64387 Seconds  
(without alignments)  
2384.688 Million cell updates/sec

Title: US-09-813-453A-15  
Perfect score: 1392  
Sequence: 1 MIIIDSGNSRLKVGWFDPD.....LDSPVLDGLAALAAQAQAPTA 267

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues  
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1392	100.0	267	1	BAF_BORPE
2	267	19.2	54	1	BAF_BORBR
3	125	9.0	233	1	YACB_BACSU
4	107.5	7.7	358	1	YM05_MYCTU
5	106.5	7.7	423	1	VG75_HSVI1
6	102.5	7.4	442	1	AROA_BORPE
7	101	7.3	1110	1	CYGD_BOVIN
8	99.5	7.1	700	1	PURL_HALN1
9	98.5	7.1	442	1	AROA_BORBR
10	98.5	7.1	676	1	ICP0_HSVBK
11	94.5	6.8	449	1	SNDH_ACELI
12	92.5	6.6	440	1	GAT4_MOUSE
13	92	6.6	267	1	THIG_DEIRA
14	91	6.5	408	1	BPHG_BURCE
15	90.5	6.5	351	1	KLF2_RAT
16	90	6.5	759	1	YEHM_ECOLI
17	89.5	6.4	442	1	GAT4_HUMAN
18	89.5	6.4	507	1	VL2_HPVI1A
19	89.5	6.4	559	1	NERA_ALCSP
20	89.5	6.4	747	1	ELS_BOVIN
21	89	6.4	1140	1	BCSC_ECOLI
22	88.5	6.4	361	1	COBT_MYCTU
23	88.5	6.4	1083	1	T2D3_HUMAN
24	88.5	6.4	3530	1	MY15_HUMAN
25	88	6.3	462	1	MURC_STRCO
26	88	6.3	530	1	MURE_MYCLE
27	87.5	6.3	354	1	ODPA_ZYMMO
28	87.5	6.3	459	1	COBG_PSEDE
29	87.5	6.3	1108	1	CYGE_RAT
30	87.5	6.3	1625	1	CTPL_MYCTU
31	87.5	6.3	2205	1	POLN_RUBVT
32	87	6.2	465	1	E2F3_HUMAN
33	87	6.2	631	1	Y282_MYCTU

34	87	6.2	655	1	PRIA_MYCTU	P71670 mycobacteri
35	86.5	6.2	226	1	BIOD_MYCTU	O06620 mycobacteri
36	86.5	6.2	750	1	ELS_CHICK	P07916 gallus gall
37	86	6.2	434	1	UL43_HSVI1	P10227 herpes simp
38	86	6.2	516	1	AROA_BRANA	P17688 brassica na
39	85.5	6.1	383	1	MRP_MYCLE	P53382 mycobacteri
40	85.5	6.1	409	1	AMBI_BACST	P37113 bacillus st
41	85.5	6.1	499	1	FEAB_ECOLI	P80668 escherichia
42	85.5	6.1	1044	1	RSG1_BOVIN	P09851 bos taurus
43	85	6.1	375	1	ACT1_TETTH	P10992 tetrahymena
44	85	6.1	376	1	ACT2_TETPY	P10993 tetrahymena
45	85	6.1	724	1	MALQ_MYCTU	O53932 mycobacteri

ALIGNMENTS

RESULT 1  
BAF\_BORPE  
ID BAF\_BORPE STANDARD; PRT; 267 AA.  
AC Q45338; Q45373;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Bvg accessory factor.  
GN BAF.  
OS Bordetella pertussis.  
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;  
OC Bordetella.  
OX NCBI\_TaxID=520;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BP504;  
RX MEDLINE=95325323; PubMed=7601846;  
RA Deshazer D., Wood G.E., Friedman R.L.;  
RT "Identification of a Bordetella pertussis regulatory factor required  
RT for transcription of the pertussis toxin operon in Escherichia  
RT coli.";  
RL J. Bacteriol. 177:3801-3807(1995).  
RN [2]  
RP SEQUENCE OF 1-38 FROM N.A.  
RC STRAIN=BP504;  
RA Wood G.E., Friedman R.L.;  
RT "Identification of a bira homolog in Bordetella pertussis.";  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 239-267 FROM N.A.  
RC STRAIN=BP536;  
RX MEDLINE=96419162; PubMed=8821935;  
RA Allen A.G., Maskell D.J.;  
RT "The identification, cloning and mutagenesis of a genetic locus  
RT required for lipopolysaccharide biosynthesis in Bordetella  
RT pertussis.";  
RL Mol. Microbiol. 19:37-52(1996).  
CC -!- FUNCTION: ACTIVATES TRANSCRIPTION OF THE PERTUSSIS TOXIN OPERON IN  
CC A BVGAS-DEPENDENT MANNER. MAY INTERACT WITH THE ALPHA SUBUNIT OF  
CC RNA POLYMERASE.  
CC  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; U12020; AAA75361.1; -.  
CC EMBL; AF016461; AAC68834.1; -.  
CC EMBL; X50711; CAA62242.1; -.  
CC InterPrc; IPR004619; Baf.  
CC Pfam; PF03309; Bvg\_acc\_factor; 1.  
KW Transcription regulation; Activator.  
SQ SEQUENCE 267 AA; 27845 MW; 336A615F67B57901 CRC64;



Matches 48; Conservative 21; Mismatches 65; Indels 22; Gaps 6;

QY 83 LRAQPLAMGLRNG-----YRNPDLGADRWACWVGVLARQPSVHPPLLVASFGTATLDTI 138  
Db 83 IEPQIVGPGMKTGLNIKNDPNKEVGADR---IVNAVAIHLGNPLIVDFGTATTCYI 139

QY 139 GPDNVFPGGLILPGPAMRGALAYGTAHLPLADGLVADYPI--DTHQAIASGIAAQAAGA 196  
Db 140 DENKQYMGGAIAPIGTTISTEALYSRAAKLPRIETRPDNIIGKNTVSAMQSGILPGYVQG 199

QY 197 IVRWLAGRQRY--GQAPEIYVAGGGWPEVRQEAER 230  
Db 200 V--EGIVKRMKWQAKQDPR-----SLRQEAWR 224

RESULT 4

YMO5\_MYCTU

ID YMO5\_MYCTU STANDARD; PRT; 358 AA.

C Q10394;  
T 01-OCT-1996 (Rel. 34, Created)  
T 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein Rv2205c.  
GN Rv2205C OR MT2261 OR MTCY190.16C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeler K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence."  
RL Nature 393:537-544(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
A Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
A Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains."  
RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
RL -I- SIMILARITY: BELONGS TO THE GLYCERATE KINASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Z70283; CAA94248.1; ALT\_INIT.  
DR EMBL; AE007072; AAK46547.1; ALT\_INIT.  
DR TIGR; MT2261; -.  
DR TubercuList; Rv2205c; -.  
DR InterPro; IPR004381; Cons\_hypoth45.  
DR InterPro; IPR003747; Glycerate\_kinase.  
DR Pfam; PF02595; DUF168; 1.  
DR TIGRFAMS; TIGR00045; Cons\_hypoth45; 2.  
KW Hypothetical protein; Transferase; Kinase; Complete proteome.

SQ SEQUENCE 358 AA; 35592 MW; AC7EDC5BF8E41544 CRC64;

Query Match 7.7%; Score 107.5; DB 1; Length 358;  
Best Local Similarity 26.5%; Pred. No. 0.28;  
Matches 75; Conservative 30; Mismatches 99; Indels 79; Gaps 15;

QY 9 NSRLKVGW-FDPDAPQAAREPAPVAFDNLDDLALGRWLATLPRRQORALGVNVAGLARGE 67  
Db 69 NTVVNAAWVDPGSATAYLECA-----QACGLGLGGPPTPETALAAHSGV--GQ 117

QY 68 AIAATLRAGGCDIRWLRAQPLAMGLRNGYRNPD-----QLG-----ADRWACWVGVLAR 116  
Db 118 LIAAALRAG-----AARIVVGL-GGSACTDGGKGMIAELGGLDAARROLADVEVIAA 168

QY 117 QPSVHPPLLVASFGTATTLDITIGPDNVFPGGLILPGPAMRGALAYGTAHLPLADGLVAD 176  
Db 159 SDVEYP--LLGPWGTA-----RVF-----APQKA---DMATVAVLEGRLLAA 205

QY 177 YPIDTHQAIASGIA-----AAQAGAIVRQWLA--GRQRYGQA-----PEIYV 216  
Db 206 WAIELDAAAGRGVSAEPGAGAGGIGAGLLAVGGRYQSGAAIIAEHTHFADDLADAELIV 265

QY 217 AGGWPEVRQEAERLLAVTGAAGFATPQPTYLDSPVLDGLAAL 259  
Db 266 TGEG----RFDEQSLHGKVVGAIAAAARPLAIPVIVLAGQVSL 304

RESULT 5

VG75\_HSVII

ID VG75\_HSVII STANDARD; PRT; 423 AA.

AC Q00129;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 01-DEC-1992 (Rel. 24, Last annotation update)  
DE Hypothetical gene 75 protein.  
GN 75.  
OS Ictalurid herpesvirus 1 (Channel catfish virus) (CCV).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Ictalurid Herpes-like viruses.  
OX NCBI\_TaxID=10401;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Auburn 1;  
RX MEDLINE=92087490; PubMed=1727613;  
RA Davison A.J.;  
RT "Channel catfish virus: a new type of herpesvirus."  
RL Virology 186:9-14(1992).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M75136; AAA88177.1; -.  
DR PIR; A36794; A36794.  
KW Hypothetical protein.  
SQ SEQUENCE 423 AA; 46531 MW; 07AC0F1D9CEE0F7 CRC64;

Query Match 7.7%; Score 106.5; DB 1; Length 423;  
Best Local Similarity 23.4%; Pred. No. 0.39;  
Matches 50; Conservative 24; Mismatches 65; Indels 75; Gaps 11;

QY 56 LGVNVAGLARGEIAATLRAGGCDIRWLRAQPLAMGLRNGY----- 96  
Db 151 MGPPVQGMPP--ETISITINPFGWFSWLVLKPAAGTGSRKPFANDLTRYLRTSGFSTAGI 208

QY 97 -----RNPDL-GADRWACWVGVLARQPSVHPPLLVASFGTATLDTIGPDNVFPGGLIL 150  
Db 209 DRFETRSPIILYGGGRW-----VLECVPHLHPTLTVAKESIECALD-----PGSIV 255



QY 151 --PGPAMRGALAYGTAHLPLADGLVADYPIDTHQAIASGIAA-----AQAGAIVR 199  
Db 256 IWPGPA-----LPVVRHVLETVLVDVRDLVGTLANFEDIFVHRVDRDRVTH 303  
QY 200 QWL-----AGRQRY-----GQAPEIYVA--GGGW 221  
Db 304 RWVEIPKNAPRKLSFTVTGHRPEVLAMCGGGW 337

RESULT 6

AROA\_BORPE STANDARD; PRT; 442 AA.  
AC P12421;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) (5-enolpyruvylshikimate-3-phosphate synthase) (EPSP synthase) (EPSPS).  
GN AROA.  
OS Bordetella pertussis.  
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;  
OC Bordetella.  
OX NCBI\_TaxID=520;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88227818; PubMed=28973356;  
RA Maskell D.J., Morrissey P., Dougan G.;  
RT "Cloning and nucleotide sequence of the aroA gene of Bordetella pertussis.";  
RL J. Bacteriol. 170:2467-2471(1988).  
CC -|- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate - phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.  
CC -|- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway; sixth step.  
CC -|- SUBUNIT: MONOMER.  
CC -|- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
CC -|- SIMILARITY: BELONGS TO THE EPSP SYNTHASE FAMILY.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; M20023; AAA22968.1; -.  
PIR; A32007; XUBRVS.  
InterPro; IPR001986; EPSP\_synthase.  
Pfam; PF00275; EPSP\_synthase; 1.  
ProDom; PD001867; EPSP\_synthase; 1.  
PROSITE; PS00104; EPSP\_SYNTHASE\_1; 1.  
PROSITE; PS00885; EPSP\_SYNTHASE\_2; 1.  
KW Aromatic amino acid biosynthesis; Transferase.  
SQ SEQUENCE 442 AA; 46688 MW; D81088B36D967955 CRC64;

Query Match 7.4%; Score 102.5; DB 1; Length 442;  
Best Local Similarity 26.2%; Pred. No. 0.82;  
Matches 68; Conservative 22; Mismatches 83; Indels 87; Gaps 13;

QY 45 LATLRRPQRALGVNVAGLAR-----GEAIAATLRAGGCDIR----- 81  
Db 115 LSGVPRMHERPIGDLVDALRQFGAGIEYLQAGYPPRLRIGGGSIRVDGPVRVEGSVSQF 174  
QY 82 ---WLRAPQPLAMGLRNGYRNPDLQAGADRWACMVGVLARQPSVHPL-LVASFGTATLTD- 136  
Db 175 LTALLMAAPV-LARRS-----GQDITIEVVGELISKPYIETLNLMARFGVSVRRDG 225  
QY 137 ----TIGPDNVF--PGGLILPGPAMRGALAYGTAHLPLADGLVADYPI-----DTHQ 183  
Db 226 WRAPTIARDAVYRGPRMAIEGDA-----STASYFLALGALGGGPVRVTGVGEDSIQ 277  
QY 184 A-IASGIAAAQAGAIVRQWLAGRQRYGQAPEIYVAGGQWPEVR-----QEAERLLA----- 233

RESULT 7

CYGD\_BOVIN STANDARD; PRT; 1110 AA.  
ID CYGD\_BOVIN  
AC P55203; O02809;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Retinal guanylyl cyclase 1 precursor (EC 4.6.1.2) (Guanylate cyclase 2D, retinal) (RETGC-1) (Rod outer segment membrane guanylate cyclase) (ROS-GC) (Guanylate cyclase E) (GC-E).  
DE GUCY2D OR GUC2D.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Retina;  
RX MEDLINE=94379976; PubMed=7916565;  
RA Goraczniak R.M., Duda T., Sitaramayya A., Sharma R.K.;  
RT "Structural and functional characterization of the rod outer segment membrane guanylate cyclase.";  
RL Biochem. J. 302:455-461(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Duda T., Venkataraman V., Sharma R.K.;  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Holstein; TISSUE=Retina;  
RX MEDLINE=97398155; PubMed=9256080;  
RA Johnston J.P., Farhangfar F., Aparicio J.G., Nam S.H., Applebury M.L.;  
RT "The bovine guanylate cyclase GC-E gene and 5' flanking region.";  
RL Gene 193:219-227(1997).  
RN [4]  
RP SEQUENCE OF 57-71; 163-178; 702-713 AND 749-763.  
RC TISSUE=Retina;  
RX MEDLINE=93343947; PubMed=8102054;  
RA Margulis A., Goraczniak R.M., Duda T., Sharma R.K., Sitaramayya A.;  
RT "Structural and biochemical identity of retinal rod outer segment membrane guanylate cyclase.";  
RL Biochem. Biophys. Res. Commun. 194:855-861(1993).  
CC -|- FUNCTION: PROBABLY PLAYS A SPECIFIC FUNCTIONAL ROLE IN THE RODS AND/OR CONES OF PHOTORECEPTORS. IT MAY BE THE ENZYME INVOLVED IN THE RESYNTHESIS OF CGMP REQUIRED FOR RECOVERY OF THE DARK STATE AFTER PHOTOTRANSDUCTION.  
CC -|- CATALYTIC ACTIVITY: GTP = 3',5'-cyclic GMP + diphosphate.  
CC -|- ENZYME REGULATION: INHIBITED BY ATP-MG OR ATP-MN.  
CC -|- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -|- TISSUE SPECIFICITY: SPECIFICALLY EXPRESSED IN THE RETINA.  
CC -|- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-4/GUANYLYL CYCLASE FAMILY.  
CC -|- SIMILARITY: CONTAINS 1 PROTEIN KINASE-LIKE DOMAIN.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; L37089; AAA50790.1; -.  
EMBL; AF027203; AAB86385.1; -.











Db 42 PRVPSSVLGLSYL-QGGGSAAGTTSGSS-----GAGPSGAGPGTQGGSPGWSQAGAE- 95  
QY 107 WACWVGVLARQPSVHPPLLVASFSGTATLTDTIGPDNVFPGG-----LILPGPAMMRGALAY 162  
Db 96 -----GAAYTPPVSPRFSPTGTTGSLAAAAAAREAAAY 132  
QY 163 GTAHLPLADGLVADYPIDTHQAIASGIAAAQAGAIVRWLACRQRYGQ-----AP-E 213  
Db 133 G-----SGGAAGAG-----LAGREQYGRPGFAGSYSSYP 163  
QY 214 IYVA--GGWPEVRQEAERLLAVTGAAGATPQPTLYLDSPLVDGLAALAAQG 263  
Db 164 AYMAVVGASW-----AAAAAASAGP--FDSPLVHSLPGRANPG 199

RESULT 13  
THIG\_DEIRA  
ID THIG\_DEIRA STANDARD; PRT; 267 AA.  
AC Q9RYX1;  
DT 16-OCT-2001 (Rel. 40, Created)  
T 16-OCT-2001 (Rel. 40, Last sequence update)  
F 16-OCT-2001 (Rel. 40, Last annotation update)  
J Thiazole biosynthesis protein thig.  
GN THIG OR DRA0172.  
OS Deinococcus radiodurans.  
OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;  
OC Deinococcaceae; Deinococcus.  
OX NCBI\_TaxID=1299;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=R1;  
RX MEDLINE=20036896; PubMed=10567266;  
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,  
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
RA Fraser C.M.;  
RT "Genome sequence of the radioresistant bacterium Deinococcus  
radiodurans R1.";  
RL Science 286:1571-1577(1999).  
CC -!- FUNCTION: Required for the synthesis of the thiazole moiety of  
CC thiamine (By similarity).  
CC -!- COFACTOR: FMN (Potential).  
CC -!- PATHWAY: Thiamine biosynthesis.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -!- SIMILARITY: BELONGS TO THE THIG FAMILY.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; AE001862; AAF12198.1; ALT\_INIT.  
DR TIGR; DRA0172;  
DR InterPro; IPR003009; FMN\_enzyme.  
KW Thiamine biosynthesis; Flavoprotein; FMN; Complete proteome.  
SQ SEQUENCE 267 AA; 27590 MW; A4ED0A2D37DAA98 CRC64;

Query Match 6.6%; Score 92; DB 1; Length 267;  
Best Local Similarity 29.0%; Pred. NO. 3;  
Matches 65; Conservative 21; Mismatches 70; Indels 68; Gaps 15;  
QY 34 DNLDLALGRWLATLP-----RRPQALGVNVAGLARGEIAATLRAGGCDIRWLAQPL 88  
Db 62 DALDLDR-----FQLLPNTAGCRTAEAA--VRVAKLAR-----AAT-----GVSWLKLEVI 105  
QY 89 AMGLRNGYRNPDLGADRWACMV-----GVLARQ-----PSVHPPLLVAS 128

Db 106 P---DPKYLLPDPIGITLRAAEILVGEGETVLPYVQPDGVLARALEAVGCATVMP--LASP 160  
QY 129 FGTA-----TTLDTIGPDNVFPGGLILPGPAMMRGALAYGTAHLPLADGLVADY- 177  
Db 161 IGTGKGLRTPELIRTVIDGAGVPIVVDAGLGVPSDAAQ--ALELG-ADAVLVNTAIEAR 217  
QY 178 -PIDTHQAIASGIAAAQAGAIVRWLACR--QRYGOAPEIYVAG 218  
Db 218 DPVAMARAFALGVEAGRLG-----YLAGRMQQRDSASPSSPAAG 256

RESULT 14  
BPHG\_BURCE  
ID BPHG\_BURCE STANDARD; PRT; 408 AA.  
AC P37337;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Biphenyl dioxygenase system ferredoxin--NAD(+) reductase component  
DE (EC 1.18.1.3).  
GN BPHG.  
OS Burkholderia cepacia (Pseudomonas cepacia).  
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;  
OC Burkholderia.  
OX NCBI\_TaxID=292;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LB400;  
RX MEDLINE=92234948; PubMed=1569021;  
RA Erickson B.D., Mondello F.J.;  
RT "Nucleotide sequencing and transcriptional mapping of the genes  
encoding biphenyl dioxygenase, a multicomponent  
polychlorinated-biphenyl-degrading enzyme in Pseudomonas strain  
LB400.";  
RL J. Bacteriol. 174:2903-2912(1992).  
RN [2]  
RP REVISIONS.  
RA Erickson B.D., Mondello F.J.;  
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: PART OF THE ELECTRON TRANSFER COMPONENT OF BIPHENYL  
CC DIOXYGENASE, TRANSFERS ELECTRONS FROM FERREDOXIN (BPHF) TO  
CC NADH.  
CC -!- CATALYTIC ACTIVITY: Reduced ferredoxin + NAD(+) = oxidized  
CC ferredoxin + NADH.  
CC -!- COFACTOR: FAD.  
CC -!- PATHWAY: Biphenyl and polychlorinated biphenyl degradation.  
CC -!- SUBUNIT: THIS DIOXYGENASE SYSTEM CONSISTS OF FOUR PROTEINS: THE  
CC TWO SUBUNITS OF THE HYDROXYLASE COMPONENT (BPHA AND BPHE), A  
CC FERREDOXIN (BPHF) AND A FERREDOXIN REDUCTASE (BPHG).  
CC -!- SIMILARITY: TO OTHER BACTERIAL RING-HYDROXYLATING DIOXYGENASE  
CC FERREDOXIN REDUCTASE COMPONENTS.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; M86348; AAB63429.1; -  
DR PIR; F41858; F41858.  
DR InterPro; IPR001327; FAD\_pyr\_redux.  
DR InterPro; IPR000205; NAD\_binding.  
DR Pfam; PF00070; pyr\_redux; 1.  
DR PRINTS; PR00368; FADPNR.  
DR ProDom; PD000139; FAD\_pyr\_redux; 1.  
KW Aromatic hydrocarbons catabolism; Flavoprotein; FAD; NAD;  
Oxidoreductase.  
FT NP\_BIND 4 35 FAD (ADP PART) (POTENTIAL).  
FT NP\_BIND 145 173 NAD (ADP PART) (POTENTIAL).  
SQ SEQUENCE 408 AA; 42953 MW; 8A52BB01688667A9 CRC64;





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:35:06 ; Search time 21.7766 Seconds  
(without alignments)  
2526.317 Million cell updates/sec

Title: US-09-813-453A-15  
Perfect score: 1392  
Sequence: 1 MIILIDSGNSRLKVGWFDPPD.....LDSPVLDGLAALAAQGAPTA 267

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTRMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	311	22.3	295	16 Q8Y2M4	ralstonia s
2	256.5	18.4	592	16 Q9JWI7	Q9jwi7 neisseria m
3	254.5	18.3	592	16 Q9JXF1	Q9jxf1 neisseria m
4	194.5	14.0	242	16 Q9PCI4	Q9pci4 xylella fas
5	178.5	12.8	248	16 Q9HWC1	Q9hwc1 pseudomonas
6	161	11.6	259	16 Q8YAC5	Q8yac5 listeria mo
7	160.5	11.5	273	16 Q97EB4	Q97eb4 clostridium
8	156	11.2	255	16 Q8R7M2	Q8r7m2 thermoanaer
9	155	11.1	259	16 Q92F54	Q92f54 listeria in
10	152	10.9	246	16 Q9WZY5	Q9wzy5 thermotoga
11	138	9.9	212	2 Q32514	Q32514 desulfovibr
12	134.5	9.7	261	16 Q9A6Z1	Q9a6z1 caulobacter
13	134	9.6	265	16 Q9X8N6	Q9x8n6 streptomyce
14	131.5	9.4	254	16 Q9KGH5	Q9kgh5 bacillus ha
15	128.5	9.2	259	16 Q8XHL5	Q8xhl5 clostridium
16	124.5	8.9	258	2 Q9F985	Q9f985 bacillus st

17	122	8.8	251	16 Q9I6F4	Q9i6f4 pseudomonas
18	112.5	8.1	340	16 Q9RJI2	Q9rji2 streptomyce
19	112	8.0	9477	2 Q9L4X3	Q9l4x3 streptomyce
20	110.5	7.9	583	2 O87808	O87808 pseudomonas
21	110.5	7.9	591	2 O69494	O69494 mycobacteri
22	110.5	7.9	600	16 Q9CD17	Q9cd17 mycobacteri
23	109	7.8	636	2 Q9X760	Q9x760 leptothrix
24	109	7.8	3352	2 Q93H83	Q93h83 streptomyce
25	107.5	7.7	561	2 O8RNW9	O8rnw9 streptomyce
26	107	7.7	4848	2 O07944	O07944 streptomyce
27	105.5	7.6	6145	2 Q93H84	Q93h84 streptomyce
28	105	7.5	1841	2 O33958	O33958 streptomyce
29	104.5	7.5	428	16 Q9F380	Q9f380 streptomyce
30	104.5	7.5	771	2 Q9ANE8	Q9ane8 bradyrhizob
31	103.5	7.4	262	16 Q9RX54	Q9rx54 deinococcus
32	103.5	7.4	10917	2 Q93NW6	Q93nw6 streptomyce
33	103	7.4	974	16 O86683	O86683 streptomyce
34	102.5	7.4	465	16 Q9AB09	Q9ab09 caulobacter
35	102.5	7.4	467	16 Q93J72	Q93j72 streptomyce
36	102.5	7.4	603	16 Q9AA58	Q9aa58 caulobacter
37	102	7.3	338	2 P72448	P72448 streptomyce
38	101.5	7.3	207	4 Q8WYX6	Q8wyx6 homo sapien
39	101.5	7.3	429	16 Q9RD83	Q9rd83 streptomyce
40	101.5	7.3	1343	12 Q06635	Q06635 bovine herp
41	101.5	7.3	3546	2 Q9F830	Q9f830 micromonosop
42	101	7.3	309	16 Q9RY83	Q9ry83 deinococcus
43	101	7.3	566	16 O70005	O70005 streptomyce
44	100.5	7.2	449	16 Q9RDE6	Q9rde6 streptomyce
45	100.5	7.2	546	16 Q9I2J7	Q9i2j7 pseudomonas

ALIGNMENTS

RESULT 1  
O8Y2M4 PRELIMINARY; PRT; 295 AA.  
AC Q8Y2M4;  
DT 01-MAR-2002 (TREMBlrel. 20, Created)  
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Putative transcription regulation accessory factor transcription  
DE regulator protein.  
GN RSC0311 OR RS03278.  
OS Ralstonia solanacearum (Pseudomonas solanacearum).  
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
OC Ralstonia.  
OX NCBI\_TaxID=305;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-GM11000;  
RX MEDLINE=21681879; PubMed=11823852;  
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,  
RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,  
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,  
RA Sigulier P., Thebault P., Whalen M., Wincker P., Levy M.,  
RA Weissenbach J., Boucher C.A.;  
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";  
RL Nature 415:497-502(2002).  
DR EMBL; AL646058; CAD13839.1; -.  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
KW Complete proteome.  
SQ SEQUENCE 295 AA; 30990 MW; 133074A7764BFCBB CRC64;

Query Match 22.3%; Score 311; DB 16; Length 295;  
Best Local Similarity 34.7%; Pred. No. 2e-14;  
Matches 103; Conservative 31; Mismatches 103; Indels 60; Gaps 11;  
QY 2 IILIDSGNSRLKVGWFDPPDAPQAAREP-----APVAFDNLDDLALGRWL 45  
Db 13 LLLIDAGNTRIKWAWTAADVAPPVAPGGTPWQHAGARPHDQLAELVEDWRDCHA-GAGM 71





```
QY 239 FGATPQPTYLDSPVLDGLA-ALAAQG 263
Db 561 FLAENTVRVADNLVIYGLLNMTAAEG 586

RESULT 4
Q9PCI4
ID Q9PCI4 PRELIMINARY; PRT; 242 AA.
AC Q9PCI4;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Hypothetical protein Xf1795.
GN Xf1795.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN [1]
P SEQUENCE FROM N.A.
RC STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Coutinho N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Colauto L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
XT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
DR EMBL; AE004001; AAF84603.1;
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 242 AA; 25658 MW; EDC5DE7C0449C1E9 CRC64;

Query Match 14.0%; Score 194.5; DB 16; Length 242;
Best Local Similarity 27.9%; Pred. No. 2.5e-06;
Matches 80; Conservative 25; Mismatches 105; Indels 77; Gaps 10;

QY 4 LIDSGNSRLKVGWFDPAQAAREPA-----PVAFDNLDLDALGRWLATLPRPQRAL 56
Db 5 LFDLGNRFKCA-----SLREGVIGPVTLPYLTETMDAFA---LQELPRG----- 47

QY 57 GNVVAGLARGEA-----IAATLRAGGCDIRWLRAQPLAMGLRNGYR 97
Db 48 --RVAYLASVAAPATHTHVLEVLKIHFEQVQVAATVAA-----CAGVRIAYA 92

QY 98 NPDQLGADRWACMVGLARQPSVHPPLLVSFGTATTLDLTIGPDNVFPGLILPGPAMMR 157
Db 93 HPERFGVDRFLALGSGY-----EGNVLVVGVTALTIDLLAANGCHLGGRIASPTLMR 147

QY 158 GALAYGTAHLPLADGLVADYPIDTHQAIASGIAAAQAGAIVRQWLAGRQRYGQAPEIYA 217
```

```
Db 148 QALHARAEQLPLSGGNYLEFAEDTEDALVSGCNGAVALIERSLYEAHQRLDQSVRLIHH 207
QY 218 GGGWPEVRQEAERLLIAVTGAAGFATPQPTYLDSPVLDGLAALAAQGA 264
Db 208 GGG-----VASLLPWLG---DVVHRPTL-----VLDGLAIWAAVAA 240

RESULT 5
Q9HWC1
ID Q9HWC1 PRELIMINARY; PRT; 248 AA.
AC Q9HWC1;
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Hypothetical protein PA4279.
GN PA4279.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
P SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004843; AAG07667.1;
DR InterPro; IPR004619; Baf.
DR InterPro; IPR001230; Prenyl_site.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 248 AA; 26758 MW; 609D37480899DA3B CRC64;

Query Match 12.8%; Score 178.5; DB 16; Length 248;
Best Local Similarity 29.6%; Pred. No. 3.5e-05;
Matches 75; Conservative 32; Mismatches 93; Indels 53; Gaps 11;

QY 1 MIILIDSGNSRLKVGWFDPAQAAREPAPVAFDNLDLDALGRWLAT---LPRPQRALG 57
Db 1 MILELDCGNSLIK--WRVIEG--AARSVAGGLAES--DDALVEQLTSQQALPVRACRLVS 54

QY 58 VN-----VAGLARGEIAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDLGADRWAC 109
Db 55 VRSEQTSQVLRLEQLFPVSALVASSGKQL-----AGVRNGYLDYQRLGLDRWLA 105

QY 110 MVGV--LARQPSVHPPLLVSFGTATTLDLTIGPDNVFPGLILPGPAMRGALAYGTAHL 167
Db 106 LVAAHHLAKK-----ACLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQRLTHTRRI 160

QY 168 PLADGLVADYPIDTHQAIAS---GIAAAQA-----GAIVRQWLAGRQRYGQAPEI 214
Db 161 RYDDA-----EARRALASLPQGOATAEAVERGCLLMRLRGFVREQYAMACELLGPDCEI 213

QY 215 YVAGGGWPEVRQE 227
Db 214 FLTGGDAELVRDE 226

RESULT 6
Q8YAC5
ID Q8YAC5 PRELIMINARY; PRT; 259 AA.
AC Q8YAC5;
DT 01-MAR-2002 (TremBLrel. 20, Created)
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
```



Matches 65; Conservative 41; Mismatches 109; Indels 36; Gaps 11;  
QY 1 MIILIDSGNSRLKVGWFDPPDAPQAAREPAPVAFD-NLDDLALGRWLATLPRRPQALGVN 59  
Db 1 MLLAFDVGNNTNIVMGVF---KGKLLHSFRISTDKNKTDEYGLVNQL-----IGN 50  
QY 60 VAGLARGEAIAA-----TLRAGGCDIRWLRAQPLAM--GLRNG-----YRNPDLGA 104  
Db 51 GISLTEIDDDVISSVVPPLMNTLQV--MSLKYFRTPKPIVVGPIKGTGINIKYDNPKEVGA 108  
QY 105 DRWACMGVGLARQPSVHPPLLVASFSGTATLTDIGPDNVFPFGLILPGPAMRGALAYGT 164  
Db 109 DR---IVNAAYELYGGPVIIVDFGTATTFCAISEKGEYLGIIAPGLMISADALFQRT 165  
QY 165 AHLPLADGLVADYPI--DTHQAIAASGIAAAQAAGAIVRQWLAGRQRYGOAPEIY-VAGGGW 221  
Db 166 AKLPKIDLTKPPTVINRNTVASMQSGIYGHVGMV--DYIVTRMKGEFAPSAYVVVATGGF 223  
-QY 222 PEVRQEAERLL 232  
.b 224 ANMIAEESKTI 234

RESULT 9  
Q92F54  
ID Q92F54 PRELIMINARY; PRT; 259 AA.  
AC Q92F54;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Hypothetical protein lin0253.  
GN LIN0253.

OS Listeria innocua.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Listeriaceae; Listeria.  
OX NCBI\_TaxID=1642;

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CLIP 11262 / SEROVAR 6A;  
RX PubMed=11679669;  
RA Glaser P., Frangeul L., BUCHRIESER C., Rusniok C., Amend A.,  
RA Baquero F., Bloeker P., Bloeker H., Brandt P., Chakraborty T.,  
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,  
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,  
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
RA Jones L.-M., Kaerst U., Krefit J., Kuhn M., Kunst F., Kurapkat G.,  
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,  
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,  
RT "Comparative genomics of Listeria species.";  
RL Science 294:849-852(2001).  
DR EMBL; AL596164; CAC95486.1; -.  
DR ListiList; LIN00253; -.  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMS; TIGR00671; baf; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 259 AA; 28227 MW; 554B03A0C0EFA64F CRC64;

Query Match 11.1%; Score 155; DB 16; Length 259;  
Best Local Similarity 25.9%; Pred. No. 0.0016;  
Matches 77; Conservative 38; Mismatches 110; Indels 72; Gaps 16;  
QY 1 MIILIDSGNSRLKVGWFDPPDAPQAAREPAPVAFDNLDDLALGRWLATLPR-RPQALGV- 58  
Db 1 MILVIDGNTNCTGVYK-----EQKLLRHRWMTTDRHRTSDELGMT 42  
QY 59 -----NVAGLARGEAIAATLRA--GGCDIRWLRAQPLAM--GLRNGYR----- 97  
Db 43 VLNFFSYANLTPSDIQIISVVPPIHMETMC-VRYFNIRPLIVGPGIKTGLNLKVD 101

QY 98 NPDQLGADRWACMGVGLARQPSVHPPLLVASFSGTATLTDIGPDNVFPFGLILPGPAMMR 157  
Db 102 NPREGISDR---IVNAVAASEEYGTPIVVVDFTATTCYIDEAGVYQSGAIPGIMIST 158  
QY 158 GALAYGTAHLPLADGLVADYPIDTHQAIA-SGIAAAQAG---AIVRQW---LAGRQRYGQ 210  
Db 159 EALYNRAAKLPRVD--IA----ESSQIICKSTVASMQAQAGIFYGFIGOCEGIIAEMKKQSN 212  
QY 211 APEIYVAGGWPPEVRQEAERLLAVTGAAGCATPQPTYLDS-PVLDTGLAALAAQGAAPT 266  
Db 213 TSPVVVATGGL--ARMITEKSSAV-----DILDPFLTCKGLELLELYRRNKPT 256

RESULT 10  
Q9WZY5  
ID Q9WZY5 PRELIMINARY; PRT; 246 AA.  
AC Q9WZY5;  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Hypothetical protein TM0883.  
GN TM0883.

OS Thermotoga maritima.  
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.  
OX NCBI\_TaxID=2336;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MSB8 / DSM 3109;  
RX MEDLINE=99287316; PubMed=10360571;  
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,  
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,  
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;  
RT "Evidence for lateral gene transfer between Archaea and Bacteria from  
genome sequence of Thermotoga maritima.";

RL Nature 399:323-329(1999).  
DR EMBL; AE001754; AAD35964.1; -.  
DR TIGR; TM0883; -.

DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMS; TIGR00671; baf; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 246 AA; 27154 MW; 9E0309AD462CF266 CRC64;

Query Match 10.9%; Score 152; DB 16; Length 246;  
Best Local Similarity 25.0%; Pred. No. 0.0025;  
Matches 62; Conservative 36; Mismatches 88; Indels 62; Gaps 10;

QY 1 MIILIDSGNS-----RLKVGWFDPPDAPQAAREPAPVAFDNLDDLALGRW 44  
Db 1 MYLLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTE-----DELFSH 42

QY 45 LATLPRRPQALGVNVAGLARGEAIAATLRAGGCDI-----RWLRAQPLAMGLRNG---- 95  
Db 43 LHPLLGDAMREI-----KGIGVASVVPQTQNTVIERFSQYFHSPIWVKAKNGCVKW 94

QY 96 -YRNPDLGADRWACMGVGLARQPSVHPPLLVASFSGTATLTDIGPDNVFPFGLILPGPA 154  
Db 95 NVKNPSEVGADR---VANVAVFVKEYGKNGIIDMGTTATTVDLV-VNGSYEGGAILPGFF 150

QY 155 MMRGALAYGTAHLPLADGLVADYPI--DTHQAIAASGIAAAQAAGAIVRQWLAGR--QRYGQ 210  
Db 151 MMVHSLFRGTAKLPLVEVKPADFVVGKDTTEENIRLGWNGSVYAL--EGIIIGRIKEVGD 208

QY 211 APEIYVAG 218  
Db 209 LPVVLTTGG 216

RESULT 11  
O32514

```
ID O32514 PRELIMINARY; PRT; 212 AA.
AC O32514;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Similar to Bacillus subtilis.
OS Desulfovibrio vulgaris (strain Miyazaki).
OC Bacteria; Proteobacteria; delta subdivision; Desulfovibrionaceae;
OC Desulfovibrio.
OX NCBI_TaxID=883;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MIYAZAKI;
RA Kitamura M., Konishi T., Kawanishi K., Ohashi K., Kishida Y.,
RA Kohno K., Akutsu H., Kumagai I., Nakaya T.;
RT "Sequence analyses of two ferredoxin genes and their flanking regions
RT from Desulfovibrio vulgaris (Miyazaki F).";
RL J. Biochem. Mol. Biol. Biophys. 2:147-154(1998).
DR EMBL; AB005550; BAA21476.1; -.
-DR InterPro; IPR004619; Baf.
R Pfam; PF03309; Bvg_acc_factor; 1.
Q SEQUENCE 212 AA; 22336 MW; 23E1789973A344D7 CRC64;

Query Match 9.9%; Score 138; DB 2; Length 212;
Best Local Similarity 27.4%; Pred. No. 0.02;
Matches 62; Conservative 34; Mismatches 78; Indels 52; Gaps 12;

QY 3 ILIDSGNSRLKVGWEDPDAPQAAREPAPVAFDNLDLALGRWLATLPRRPQALGVNV-- 60
   :| | | :| | :| | :| | :| | :| | :| | :| | :| | :| | :
Db 6 LLEDIGNTVNKIG-----IAVETAVLTSY--VLPTDPGQTTDSIGLRLE 48

QY 61 ----AGLAR---GEAIAATLRAG-----GCDIRWLRAQ-----PLAMGLRNGYRNP 100
   ||| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :
Db 49 VLRHAGLGPADVGACVASSVPGVGNPLIRRAE-RYLKRLLFAPGDIAIPLDNRVERPA 107

QY 101 QLGADRWACMGVGLARQPSVHPPLLLVA-SFGTATLDTIGPDNVFPGGLILPGPAMM 159
   :| | | :| | :| | :| | :| | :| | :| | :| | :| | :| | :
Db 108 EVGADR--LVAAYAAARLYPPGPRSLVSVDFGTATFTDCV-EGGAYLGGILICPGVLS 164

QY 160 LAYGTAHLP-LADGLVADYPIDTHQAIAS-----GIAAAQAGAI 197
   | :| | | :| | :| | :| | :| | :| | :| | :| | :| | :
Db 165 LSSRTAKLPRISLEVEEDSPVIGRSTTTSLNHGFIFGFAAMTEGVL 210

RESULT 12
Q9A6Z1 PRELIMINARY; PRT; 261 AA.
AC Q9A6Z1;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DE Transcriptional activator, putative, Baf family.
GN CC1935.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005867; AAK23910.1; -.
DR TIGR; CC1935; -.
DR InterPro; IPR004619; Baf.
```

```
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00671; baf; 1.
KW Complete proteome.
SQ SEQUENCE 261 AA; 27965 MW; C19E50D7B0714EF5 CRC64;

Query Match 9.7%; Score 134.5; DB 16; Length 261;
Best Local Similarity 28.8%; Pred. No. 0.046;
Matches 53; Conservative 18; Mismatches 78; Indels 35; Gaps 8;

QY 81 RWLRAQPLA-----MGLRNGYRNPDLQGLADRWACMGVGLARQPSVHPPLLLVASFGTATT 134
   | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :
Db 80 RYFNVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAMVYPG---PLWVIDSGTATT 136

QY 135 LDTIGPDNVFPGGLILPGPAMMRGALAYGTAHLPLADGLVADYPIDTHQAIASGIAAAQA 194
   | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :
Db 137 FDIWAADGAFEGGIIAPGINLSMQALHEAAKLP---RIAIQRPAGNRIVGTDTVSAMQS 193

QY 195 GAIVRQWLAGRQRYGQAPRIYVAGGGWPEVRQEAER-----LLAVTGAAF---GATPQPT 246
   | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :
Db 194 GVF---W-----GYISLIEGLVA-----RIKAERGEPTVIATCGVASLFEATDSID 238

QY 247 YLDS 250
   :| |
Db 239 HFDS 242

RESULT 13
Q9X8N6 PRELIMINARY; PRT; 265 AA.
AC Q9X8N6;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical protein SCO3380.
GN SCO3380 OR SCE94.31C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=843436;
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
```



```
DR EMBL; AL049628; CAB40880.1; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00671; baf; 1.
KW Hypothetical protein.
SQ SEQUENCE 265 AA; 28203 MW; E34FA2ABBA946969 CRC64;

Query Match          9.6%; Score 134; DB 16; Length 265;
Best Local Similarity 25.8%; Pred. No. 0.051;
Matches 59; Conservative 31; Mismatches 83; Indels 56; Gaps 8;

QY 1 MILIDSGNSRLKVGWFDPAQQAAREPAPVAFDNLDLALGRW-LATLPRRPQALGVN 59
Db 1 MLITIDVGNTHTVLGLFDGE-----DIVEHWRISTDSRRTADELAVL 42
QY 60 VAGL-----ARGEAIAATL-----RAGGCDIRWLRAQPLAMGLRN 94
Db 43 LQGLMGHPLGLDELGDGIDGIAICATVPSVLHELREVTTRYGDPVAVLVEP---GVKT 99
Y 95 GY----RNPDLGADRWACMGVGLARQPSVHPPLLVASFGTATTLDLTIGPDNVFPGLLIL 150
Db 100 GVPILTDPKVEGADR---IINAVAAVELYGGPAIVVDFTATTFFDAVSARGEYIGGVIA 156
QY 151 PGPAMMRGALAYGTAHLPLADGLIVADYPI--DTHQAIASGIAAAQAGAI 197
Db 157 PGIEISVEALGVKGAQLRKIEVARPRSVIGKNTVEAMQSGIVYGFAGQV 205

RESULT 14
Q9KGH5 PRELIMINARY; PRT; 254 AA.
AC Q9KGH5;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein BH0086.
GN BH0086.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RA "Complete genome sequence of the alkaliphilic bacterium Bacillus
.T halodurans and genomic sequence comparison with Bacillus subtilis.";
RT Nucleic Acids Res. 28:4317-4331(2000).
RL EMBL; AP001507; BAB03805.1; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 254 AA; 27907 MW; 0E3B557BA7DAC176 CRC64;

Query Match          9.4%; Score 131.5; DB 16; Length 254;
Best Local Similarity 24.1%; Pred. No. 0.072;
Matches 63; Conservative 43; Mismatches 98; Indels 57; Gaps 12;

QY 1 MILIDSGNSRLKVGWFDPAQQAAREPAPVAFDNLDLALGRW-LATLPRRPQALGVN 59
Db 1 MILVIDGNTNTVLGVYQ-----DETLVHHWRLATSRQKTEDEYAMT 42
QY 60 V-----AGL---ARGEAIAATLRAGGCDI-----RWLRAQPLAM--GLRNG----YRN 98
Db 43 VRSLFDHAGLQFQDIDGIVISSVPPMMFSLQMKCKKYFHYTPMIIGPGIKTGLNIKYN 102
QY 99 PDQLGADRWACMGVGLARQPSVHPPLLVASFGTATTLDLTIGPDNVFPGLLILPGPAMMRG 158
Db 103 PKEVGADR---IVNAVAAIELGYPAIVVDFTATTTCYCLINEKKQYAGGVIAPGIMISTE 159
```

```
QY 159 ALAYGTAHLPLAD-----GLVADYPIDTHQAIASGI---AAAQAGAIVRWLAGRQRYGQ 210
Db 160 ALYHRASKLPRIEIAKPKQVVGNTNTIDSMQ---SGIFYGVVSQDGVVKKRMKAQAE---S 213
QY 211 APEIYVAGGGWPEVRQEAERL 231
Db 214 EPKVATGGGLAKLIGTESETI 234

RESULT 15
Q8XHL5 PRELIMINARY; PRT; 259 AA.
AC Q8XHL5;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein CPE2468.
GN CPE2468.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / TYPE A;
RX PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR EMBL; AP003194; BAB82174.1; -.
DR InterPro; IPR004619; Baf.
DR InterPro; IPR000515; BPD_transp.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00671; baf; 1.
DR PROSITE; PS00402; BPD_TRANSF_INN_MEMBR; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 259 AA; 28819 MW; 0D5FEA3B7A145E10 CRC64;
```

```
Query Match          9.2%; Score 128.5; DB 16; Length 259;
Best Local Similarity 23.3%; Pred. No. 0.12;
Matches 61; Conservative 43; Mismatches 119; Indels 39; Gaps 11;

QY 1 MILIDSGNSRLKVGWFDPD--APQAAAREPAPVAFDNLDLALGRWLATLPRRPQALG 57
Db 1 MILLIDVGNNTNIVLGIHDNEKYIASWRISTDSKKTSDSEYSIQVM-QLFNQAKLPEDVEG 59
QY 58 VNVAGLARG--EAIAATLRAGGCDIRWLRAQPLAM--GLRNG----YRNPDLGADRWAC 109
Db 60 IISVVVFNIMHSLENMVRKCFC-----KEPIVVGPGIKTGINIKYDNPKEVGADR--- 110
QY 110 MGVLARQPSVHPPLLVASFGTATTLDLTIGPDNVFPGLLILPGPAMMRGALAYGTAHLPL 169
Db 111 IVNAVAAFEKHKKPMIIDIIFTATTTCATTEKGDYLGGNICPGIQISADALFERAAKLPR 170
QY 170 AD-----GLVADYPIDTHQAIASGIAAAQAGA---IVRWLAGRQRYGQAPEIYVAGGGW 221
Db 171 IELEKPKSVICKNTVTSMQA---GIIFYGKVEYIVKRMKKEMMDLGEKEPFVLATGGL 227
QY 222 -----PEVRQEAERLLAVTG 236
Db 228 AKLVSETDVIDEVDKRLTLEG 249
```

Search completed: June 24, 2003, 21:59:34  
Job time : 23.7766 secs

**THIS PAGE BLANK (USPTO)**



Db 744 SGAEAAVAAAQLAGRRTRRLRVAAHAFHSPMDGMLAGFREVAAGLRYPELTVVS 803  
QY 98 -----NPDQL-GADRWACMVGLARQP-----SVHPPLLVSFGTATTLDTIGPD 141  
Db 804 TVTGRPARPGELTGPDIWVAQV-----REPVRFADAVRTAH-----RLGARTFLET-GPD 852  
QY 142 NVF-----PGGLILPGPAMMRGALAYGTAHLPLADGLVAD 176  
Db 853 GVLGMAEECLEDDTVALLPAHKPGTAPHGPAAPG-ALRAAAAAYGRGARVDWAGMHAD 911  
QY 177 YPIDTHQAIASGIAAAQAGAIVRQWLA-GR-----QRYGOAPEIYVAGGG-----WP 222  
Db 912 GPEGPARRVLPVHAFRHR---RYWLPAGRAADTDDWYRIGWDRLPVAVTGGARTAGRWL 968  
QY 223 EVRQEAERLLAVTG-----AAFGATPQPTYLDSPVLD--GLAAL--AAQGAPT 266  
Db 969 VIHPPSPRCRELSGHAERALRAAGASPVPLVDAPADRASFAALLRSATGPDT 1022

RESULT 2

S-09-105-537-2  
; Sequence 2, Application US/09105537A  
; Patent No. 6265202  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.438US1  
; CURRENT APPLICATION NUMBER: US/09/105,537A  
; CURRENT FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 5215  
; TYPE: PRT  
; ORGANISM: Streptomyces venezuelae  
US-09-105-537-2

Query Match 7.5%; Score 105; DB 4; Length 5215;  
Best Local Similarity 24.6%; Pred. No. 0.14;  
Matches 87; Conservative 24; Mismatches 105; Indels 138; Gaps 20;  
QY 36 LDLD-----ALGRWLATLP-----RRPQAL-----GVNV 60  
Db 3473 LDLDACALVAARGRLMQLPPGGAMSVRAGEDVALLAGREDAVCVAAVNGPRSVI 3532  
61 AGLARGEATAATLRAG-GCDIRWLA-----QPLAMGLRNGYR----- 97  
Db 3533 SGAEAAVAAAQLAGRRTRRLRVAAHAFHSPMDGMLAGFREVAAGLRYPELTVVS 3592  
QY 98 -----NPDQL-GADRWACMVGLARQP-----SVHPPLLVSFGTATTLDTIGPD 141  
Db 3593 TVTGRPARPGELTGPDIWVAQV-----REPVRFADAVRTAH-----RLGARTFLET-GPD 3641  
QY 142 NVF-----PGGLILPGPAMMRGALAYGTAHLPLADGLVAD 176  
Db 3642 GVLGMAEECLEDDTVALLPAHKPGTAPHGPAAPG-ALRAAAAAYGRGARVDWAGMHAD 3700  
QY 177 YPIDTHQAIASGIAAAQAGAIVRQWLA-GR-----QRYGOAPEIYVAGGG-----WP 222  
Db 3701 GPEGPARRVLPVHAFRHR---RYWLPAGRAADTDDWYRIGWDRLPVAVTGGARTAGRWL 3757  
QY 223 EVRQEAERLLAVTG-----AAFGATPQPTYLDSPVLD--GLAAL--AAQGAPT 266  
Db 3758 VIHPPSPRCRELSGHAERALRAAGASPVPLVDAPADRASFAALLRSATGPDT 3811

RESULT 3

5310667-9  
; Patent No. 5310667

; APPLICANT: EICHHOLTZ, DAVID A.; GASSER, CHARLES S.; KISHORE,  
; GANESH M.  
; TITLE OF INVENTION: GLYPHOSATE-TOLERANT 5-ENOLPYRUVYL  
; -3-PHOSPHOSHIKIMATE SYNTHASES  
; NUMBER OF SEQUENCES: 37  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/380,963  
; FILING DATE: 17-JUL-1989  
; SEQ ID NO:9:  
; LENGTH: 440  
5310667-9  
Query Match 7.4%; Score 102.5; DB 6; Length 440;  
Best Local Similarity 26.2%; Pred. No. 0.0074;  
Matches 68; Conservative 22; Mismatches 83; Indels 87; Gaps 13;  
QY 45 LATLPRRPQRALGVNVAGLAR-----GEAIAATLRAGGCDIR----- 81  
Db 115 LSGVPRMHERPIGDLVDALRQFNAGIEYLGQAGYPPLRIGGGSIRVDGPVRVEGSSVSQF 174  
QY 82 ---WLRAOPLAMGLRNGYRNPDLQGLADRWACMVGLARQPSVHPPL-LVASFGTATTLTLD- 136  
Db 175 LTALLMAAPV-LARRS-----GQDITIEVVUGELISKPYIEITLNLMARFGVSVRRDG 225  
QY 137 ---TIGPDNVF--PGGLILPGPAMMRGALAYGTAHLPLADGLVADYPI-----DTHQ 183  
Db 226 WRAFTIARDAVYRGPRMAIEGDA-----STASYFLALGAIGGPPVVRTGVGEDSIQ 277  
QY 184 A-IASGIAAAQAGAIVRQWLAGRQRYGOAPEIYVAGGWPVEVR-----QEAERLLA----- 233  
Db 278 GDVAFATAATLAAMGADVRY-----GPGWIETRGVRVAEGGRLKAFDADF 320  
QY 234 --VTGAAFGATPQPTYLDSP 251  
Db 321 NLIPDAAMTAATLALYADGP 340

RESULT 4

US-08-476-008-64  
; Sequence 64, Application US/08476008  
; Patent No. 5627061  
; GENERAL INFORMATION:  
; APPLICANT: Barry, Gerard F.  
; APPLICANT: Kishore, Ganesh M.  
; APPLICANT: Padgett, Stephen R.  
; APPLICANT: Stallings, William C.  
; TITLE OF INVENTION: Glyphosate Tolerant  
; TITLE OF INVENTION: 5-Enolpyruvylshikimate-3-Phosphate Synthases  
; NUMBER OF SEQUENCES: 69  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dennis R. Hoerner, Jr., Monsanto Co. BB4F  
; STREET: 700 Chesterfield Village Parkway  
; CITY: St. Louis  
; STATE: Missouri  
; COUNTRY: USA  
; ZIP: 63198  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/476,008  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/306,063  
; FILING DATE: 13-SEP-1994  
; APPLICATION NUMBER: US 07/749,611  
; FILING DATE: 28-AUG-1991  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/576,537



;; FILING DATE: 31-AUG-1990  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Hoerner Jr., Dennis R.  
;; REGISTRATION NUMBER: 30,914  
;; REFERENCE/DOCKET NUMBER: 38-21(10660)A  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (314)537-6099  
;; TELEFAX: (314)537-6047  
;; INFORMATION FOR SEQ ID NO: 64:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 442 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-476-008-64

Query Match 7.4%; Score 102.5; DB 1; Length 442;  
Best Local Similarity 26.2%; Pred. No. 0.0075;  
Matches 68; Conservative 22; Mismatches 83; Indels 87; Gaps 13;  
QY 45 LATLPRRQPRALGVNVAGLAR-----GEAIAATLRAGGCDIR----- 81  
Db 115 LSGVPRMHERPIGDLVDALRQFGAGIEYLGQAGYPPPLRIGGGSIRVDGPVRVEGVSQQF 174  
QY 82 ---WLRAQPLAMGLRNGYRNPDLQGADRWACMVGLARQPSVHPPL-LVASFGTATLTD- 136  
Db 175 LTALLMAAPV-LARRS-----GQDITIEVVGELISKPYIEITLNLMAFGVSRRDG 225  
QY 137 ----TIGPDNVF--PGGLILPGPAMMRGALAYGTAHLPLADGLVADYPI-----DTHQ 183  
Db 226 WRAFTIARDAVYRGPGRMAIEGDA-----STASYFLALGAIGGGPVRVTGVEDSIQ 277  
QY 184 A-IASGIAAAQAGAIVRQWLGRQRYGQAPYIYVAGGWPVEVR-----QEAERLLA----- 233  
Db 278 GDVAFATLAAMGADVRY-----GPGWIETRGVRVAEGGRLKAFDAF 320  
QY 234 --VTGAAGATPQPTYLDSP 251  
Db 321 NLIPDAAMTAATLALYADGP 340

RESULT 5  
US-08-306-063-64  
; Sequence 64, Application US/08306063  
; Patent No. 5633435  
; GENERAL INFORMATION:  
; APPLICANT: Barry, Gerard F.  
; APPLICANT: Kishore, Ganesh M.  
; APPLICANT: Padgett, Stephen R.  
; APPLICANT: Stallings, William C.  
; TITLE OF INVENTION: Glyphosate Tolerant  
; TITLE OF INVENTION: 5-Enolpyruvylshikimate-3-Phosphate Synthases  
; NUMBER OF SEQUENCES: 69  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dennis R. Hoerner, Jr., Monsanto Co. BB4F  
; STREET: 700 Chesterfield Village Parkway  
; CITY: St. Louis  
; STATE: Missouri  
; COUNTRY: USA  
; ZIP: 63198  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/306,063  
; FILING DATE: 13-SEP-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/749,611  
; FILING DATE: 28-AUG-1991

;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/576,537  
;; FILING DATE: 31-AUG-1990  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Hoerner Jr., Dennis R.  
;; REGISTRATION NUMBER: 30,914  
;; REFERENCE/DOCKET NUMBER: 38-21(10660)A  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (314)537-6099  
;; TELEFAX: (314)537-6047  
;; INFORMATION FOR SEQ ID NO: 64:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 442 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-306-063-64

Query Match 7.4%; Score 102.5; DB 1; Length 442;  
Best Local Similarity 26.2%; Pred. No. 0.0075;  
Matches 68; Conservative 22; Mismatches 83; Indels 87; Gaps 13;  
QY 45 LATLPRRQPRALGVNVAGLAR-----GEAIAATLRAGGCDIR----- 81  
Db 115 LSGVPRMHERPIGDLVDALRQFGAGIEYLGQAGYPPPLRIGGGSIRVDGPVRVEGVSQQF 174  
QY 82 ---WLRAQPLAMGLRNGYRNPDLQGADRWACMVGLARQPSVHPPL-LVASFGTATLTD- 136  
Db 175 LTALLMAAPV-LARRS-----GQDITIEVVGELISKPYIEITLNLMAFGVSRRDG 225  
QY 137 ----TIGPDNVF--PGGLILPGPAMMRGALAYGTAHLPLADGLVADYPI-----DTHQ 183  
Db 226 WRAFTIARDAVYRGPGRMAIEGDA-----STASYFLALGAIGGGPVRVTGVEDSIQ 277  
QY 184 A-IASGIAAAQAGAIVRQWLGRQRYGQAPYIYVAGGWPVEVR-----QEAERLLA----- 233  
Db 278 GDVAFATLAAMGADVRY-----GPGWIETRGVRVAEGGRLKAFDAF 320  
QY 234 --VTGAAGATPQPTYLDSP 251  
Db 321 NLIPDAAMTAATLALYADGP 340

RESULT 6  
US-08-833-485-64  
; Sequence 64, Application US/08833485  
; Patent No. 5804425  
; GENERAL INFORMATION:  
; APPLICANT: Barry, Gerard F.  
; APPLICANT: Kishore, Ganesh M.  
; APPLICANT: Padgett, Stephen R.  
; APPLICANT: Stallings, William C.  
; TITLE OF INVENTION: Glyphosate Tolerant  
; TITLE OF INVENTION: 5-Enolpyruvylshikimate-3-Phosphate Synthases  
; NUMBER OF SEQUENCES: 69  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dennis R. Hoerner, Jr., Monsanto Co. BB4F  
; STREET: 700 Chesterfield Village Parkway  
; CITY: St. Louis  
; STATE: Missouri  
; COUNTRY: USA  
; ZIP: 63198  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/833,485  
; FILING DATE: 07-APR-1997  
; CLASSIFICATION: 435



STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII(DOS) Text only  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,227C  
FILING DATE: February 21, 1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Plant, Thomas, G.  
REGISTRATION NUMBER: 35,784  
REFERENCE/DOCKET NUMBER: X-8231  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-2459  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1996 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-804-227C-9

Query Match 7.4%; Score 102.5; DB 2; Length 1996;  
Best Local Similarity 23.8%; Pred. No. 0.066;  
Matches 94; Conservative 21; Mismatches 117; Indels 163; Gaps 18;  
QY 4 LIDSGNSR--LKVGWFDPDAP-----QAAREPAPVAFDNLDDLALGRWLATLP-- 49  
Db 1544 LVDRHKVRHLLVGRRGPDAPGVDRLVLAELTESGAEVAVRACDVTDRLALRLLDALPDE 1603  
QY 50 -----RRPQRALGVNVVAGLARG-----EAIAA 71  
Db 1604 HPLTCVVHTAGVLDGVLQAQTAERIDTVLRPKADAAVHLDLTREIGRVPLVLYSSVSA 1663  
QY 72 TLRAGGCDIRWLRAQPLAMGLRNGYRNP-----DQLGADRWACMVGVLARQPSVHPPLV- 126  
Db 1664 TLGSAG-----QAGYAAANAFMDALAARR--CAAG-----HPALSGL 1698  
QY 127 -----ASFGTATTLDITGPD-----NVFPGGLILPGPAMMRGAL----- 160  
Db 1699 WGWWSGVGLATGLD--GADAAVRRRSGLAPLDAGAALDLDLDRALTRPEPALLPVRLDLRA 1756  
QY 161 AYGTAHLPLADGLVADYPIDTHQ-----AIASGIAAAQAAGAIVRQWLAGRQR 207  
Db 1757 AAGATALPEVLRDLAGVPADARSTPGAAAGTGDEDEGAVRPAPAPADAAGTLAARLAGR-- 1814  
QY 208 YGOAPE-----IYVAGGGWPEVRQEA-----ERLLAVTGA 237  
Db 1815 --SAPERTALLDLVRTEVAAVLGHGDPAAIGAARTFKDAGFDSLTAVDLNRNLNTRTGL 1872  
QY 238 AFGAT-----POPTYLDSPVLDGLAALAAQAGPTA 267  
Db 1873 RLPATLVFDHPTPLALAEALLDGLA-AGPAEPAA 1906

RESULT 9  
US-08-804-198-3  
; Sequence 3, Application US/08804198  
; Patent No. 5945320  
; GENERAL INFORMATION:  
; APPLICANT: Burgett, Stanley G.  
; APPLICANT: Kuhstoss, Stuart A.  
; APPLICANT: Rao, Nagaraja R.  
; APPLICANT: Richardson, Mark A.  
; APPLICANT: Rostock, Paul R., Jr.  
; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PAUL R. CANTRELL 1138

STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Macintosh  
OPERATING SYSTEM: Macintosh 7.0  
SOFTWARE: Microsoft Word 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,198  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: CANTRELL, PAUL R.  
REGISTRATION NUMBER: 36,470  
REFERENCE/DOCKET NUMBER: P9113  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-3885  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1996 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-804-198-3

Query Match 7.4%; Score 102.5; DB 2; Length 1996;  
Best Local Similarity 23.8%; Pred. No. 0.066;  
Matches 94; Conservative 21; Mismatches 117; Indels 163; Gaps 18;  
QY 4 LIDSGNSR--LKVGWFDPDAP-----QAAREPAPVAFDNLDDLALGRWLATLP-- 49  
Db 1544 LVDRHKVRHLLVGRRGPDAPGVDRLVLAELTESGAEVAVRACDVTDRLALRLLDALPDE 1603  
QY 50 -----RRPQRALGVNVVAGLARG-----EAIAA 71  
Db 1604 HPLTCVVHTAGVLDGVLQAQTAERIDTVLRPKADAAVHLDLTREIGRVPLVLYSSVSA 1663  
QY 72 TLRAGGCDIRWLRAQPLAMGLRNGYRNP-----DQLGADRWACMVGVLARQPSVHPPLV- 126  
Db 1664 TLGSAG-----QAGYAAANAFMDALAARR--CAAG-----HPALSGL 1698  
QY 127 -----ASFGTATTLDITGPD-----NVFPGGLILPGPAMMRGAL----- 160  
Db 1699 WGWWSGVGLATGLD--GADAAVRRRSGLAPLDAGAALDLDLDRALTRPEPALLPVRLDLRA 1756  
QY 161 AYGTAHLPLADGLVADYPIDTHQ-----AIASGIAAAQAAGAIVRQWLAGRQR 207  
Db 1757 AAGATALPEVLRDLAGVPADARSTPGAAAGTGDEDEGAVRPAPAPADAAGTLAARLAGR-- 1814  
QY 203 YGOAPE-----IYVAGGGWPEVRQEA-----ERLLAVTGA 237  
Db 1815 --SAPERTALLDLVRTEVAAVLGHGDPAAIGAARTFKDAGFDSLTAVDLNRNLNTRTGL 1872  
QY 238 AFGAT-----POPTYLDSPVLDGLAALAAQAGPTA 267  
Db 1873 RLPATLVFDHPTPLALAEALLDGLA-AGPAEPAA 1906

RESULT 10  
US-08-804-227C-2  
; Sequence 2, Application US/08804227C  
; Patent No. 5876991  
; GENERAL INFORMATION:  
; APPLICANT: DeHoff, Bradley S.  
; APPLICANT: Kuhstoss, Stuart A.  
; APPLICANT: Rostock, Paul R., Jr.  
; APPLICANT: Sutton, Kimberly L.  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:







Db 522 XAAATYRS-----RWARKSHLLPGATGSXESCSRHPXAGGLRXACPLXACFPXPNPPE 576  
QY 110 -----MVGVLARPSV-----HPP-----LLVASFG 130  
Db 577 GARRPAIARLQKQSESTVXDRPPXACFXAFDQAVPGGPTARSPALRHLDPAYLLVSRVG 636  
QY 131 TA-----TTLDTIGPDNVFP-----GGLILPGPAMRGALAYGTA---HLP- 168  
Db 637 AAFVROXSVLVQTGGCHVFPRAQGPQRSLGHGEGVRSGLRAGQGALEAGPAGAEALPV 696  
QY 169 --LADGLVA-DYPIDTH--QAIASGIAAQAAGAIVRWL-----AGROY----- 208  
Db 697 FALRVGQVAFDHGVIVQPLRGVALGAQLALGGGAARGAVQTFEGVELGREKYRFRGVGIR 756  
QY 209 --GOAPEI--YVAGGGWP-EVROEAERLL--AVTGAAF-----GATPQPTY-LDS 250  
Db 757 AAGPADGLAFHEPGEIWPFGVKQVSPMLFADFLTSGFHEPVSTLGDERAIVRPVYRLER 816  
QY 251 PVL D 254  
Db 817 PVLE 820

RESULT 15  
US-08-804-227C-3  
; Sequence 3, Application US/08804227C  
; Patent No. 5876991  
; GENERAL INFORMATION:  
; APPLICANT: DeHoff, Bradley S.  
; APPLICANT: Kuhstoss, Stuart A.  
; APPLICANT: Rostock, Paul R., Jr.  
; APPLICANT: Sutton, Kimberly L.  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: THOMAS G. PLANT 1501  
; STREET: LILLY CORPORATE CENTER  
; CITY: INDIANAPOLIS  
; STATE: IN  
; COUNTRY: USA  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: ASCII(DOS) Text only  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/804,227C  
; FILING DATE: February 21, 1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plant, Thomas, G.  
; REGISTRATION NUMBER: 35,784  
; REFERENCE/DOCKET NUMBER: X-8231  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-2459  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1864 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-804-227C-3

Query Match 6.5%; Score 91; DB 2; Length 1864;  
Best Local Similarity 25.0%; Pred. No. 1.1;  
Matches 78; Conservative 29; Mismatches 123; Indels 82; Gaps 17;  
QY 2 IILIDSGNSRLKVGWFDPDAPQAAREPAPVAFDNLDLALGR---WLATLPRRQAL-- 56  
Db 1022 LVLPEDEGEVRLRVVLAEPDA-SGVRE-----LSVHSAGEDGGW-----TRHATAVLDT 1068  
QY 57 GNVNAGLARGEIAATLRAGGCDIRWLRAQ-----PLAMGLRNGYRNPDLGA 104

Db 1069 GTTTAGEPAG-APPAAWPPGDAEPLDLAAEYERFADAGIGYGPFAFRGLRSAWRDGDAILA 1127  
QY 105 DRWACMVGVLARQP---SVHPPLLVASFGTATTLDITIGPDNVFP---GGLILPGPAMMRG 158  
Db 1128 D--VRLPGELAGEADRYGIHPALLDAALHTAAAA-LGGAHGMPLPFTWNGVTLHA-----RG 1180  
QY 159 ALAYGTAHLPLADGLVADYPID-----THQAIASG-IAAAQAAGAIVRQWLGR 205  
Db 1181 AHAI RVLTPAGPDVA VAVTA VDPAGRPVFTAASLT LRPVT TGTQLTAAEA----- 1229  
QY 206 QRYGOAPEIYVAGGGWPEVRQEAERLLAVTGAAF-----GATPQPTYLDSPVL D G----- 255  
Db 1230 ---ARAPLYRVRWTGLPDTGTARDHTWAVAGGPGDLLPGETPHHPDLASALADTGTAPER 1286  
QY 256 -LAALAAQGAPT 266  
Db 1287 VLADLRGYGTAT 1298

Search completed: June 24, 2003, 21:49:31  
Job time : 9.7999 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:08:46 ; Search time 12.1733 Seconds  
(without alignments)  
2373.336 Million cell updates/sec

Title: US-09-813-453A-15  
Perfect score: 1392  
Sequence: 1 MIILIDSGNSRLKVGWFDPD.....LDSPLVDGLAALAAQGAPTA 267

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues  
otal number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1392	100.0	267	9 US-09-813-453A-15	Sequence 15, Appl
2	258.5	18.6	460	9 US-09-813-453A-39	Sequence 39, Appl
3	256.5	18.4	592	9 US-09-813-453A-22	Sequence 22, Appl
4	254.5	18.3	592	9 US-09-813-453A-43	Sequence 43, Appl
5	194.5	14.0	242	9 US-09-813-453A-65	Sequence 65, Appl
6	178.5	12.8	248	9 US-09-813-453A-20	Sequence 20, Appl
7	163	11.7	244	9 US-09-813-453A-41	Sequence 41, Appl
8	163	11.7	249	9 US-09-813-453A-61	Sequence 61, Appl
9	154.5	11.1	241	9 US-09-813-453A-63	Sequence 63, Appl
10	152	10.9	246	9 US-09-813-453A-9	Sequence 9, Appl
11	148	10.6	255	9 US-09-813-453A-7	Sequence 7, Appl
12	141	10.1	250	9 US-09-813-453A-3	Sequence 3, Appl
13	140.5	10.1	258	9 US-09-813-453A-49	Sequence 49, Appl
14	138	9.9	212	9 US-09-813-453A-59	Sequence 59, Appl
15	138	9.9	262	9 US-09-813-453A-45	Sequence 45, Appl
16	136.5	9.8	257	9 US-09-813-453A-53	Sequence 53, Appl
17	134.5	9.7	260	9 US-09-813-453A-51	Sequence 51, Appl
18	134	9.6	258	9 US-09-813-453A-6	Sequence 6, Appl
19	134	9.6	265	9 US-09-813-453A-4	Sequence 4, Appl

20	132	9.5	249	9 US-09-813-453A-70	Sequence 70, Appl
21	131.5	9.4	254	9 US-09-813-453A-47	Sequence 47, Appl
22	130.5	9.4	219	9 US-09-813-453A-57	Sequence 57, Appl
23	129.5	9.3	258	9 US-09-813-453A-2	Sequence 2, Appl
24	127.5	9.2	256	9 US-09-813-453A-55	Sequence 55, Appl
25	125	9.0	233	9 US-09-813-453A-17	Sequence 17, Appl
26	105	7.5	5215	9 US-09-860-846-2	Sequence 2, Appl
27	105	7.5	5215	9 US-09-988-384B-2	Sequence 2, Appl
28	105	7.5	5215	9 US-09-836-821-2	Sequence 2, Appl
29	105	7.5	5215	10 US-09-861-289-2	Sequence 2, Appl
30	103.5	7.4	262	9 US-09-813-453A-8	Sequence 8, Appl
31	102.5	7.4	442	9 US-09-464-099A-64	Sequence 64, Appl
32	102.5	7.4	442	9 US-10-214-766-36	Sequence 36, Appl
33	102.5	7.4	442	10 US-09-861-696-64	Sequence 64, Appl
34	97	7.0	257	9 US-09-813-453A-13	Sequence 13, Appl
35	97	7.0	562	9 US-09-712-363-254	Sequence 254, App
36	95	6.8	408	9 US-09-738-626-5136	Sequence 5136, Ap
37	94.5	6.8	437	9 US-10-145-415-101	Sequence 101, App
38	92.5	6.6	272	9 US-09-712-363-276	Sequence 276, App
39	92.5	6.6	272	9 US-09-813-453A-5	Sequence 5, Appl
40	90.5	6.5	891	9 US-09-976-059-18	Sequence 18, Appl
41	89.5	6.4	259	10 US-09-820-893-87	Sequence 87, Appl
42	89.5	6.4	442	10 US-09-749-728B-11	Sequence 11, Appl
43	89	6.4	4150	9 US-09-808-880-2	Sequence 2, Appl
44	88.5	6.4	190	10 US-09-764-864-1071	Sequence 1071, Ap
45	88.5	6.4	1023	9 US-09-893-519A-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1  
US-09-813-453A-15  
; Sequence 15, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 15  
; LENGTH: 267  
; TYPE: PRT  
; ORGANISM: Bordetella pertussis  
US-09-813-453A-15

Query Match	100.0%;	Score 1392;	DB 9;	Length 267;
Best Local Similarity	100.0%;	Pred. No. 3.2e-114;		
Matches 267;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
QY	1	MIILIDSGNSRLKVGWFDPDPAQAAREPAPVAFDNLDLALGRWLATLPRRPQALGVNV	60	
Db	1	MIILIDSGNSRLKVGWFDPDPAQAAREPAPVAFDNLDLALGRWLATLPRRPQALGVNV	60	
QY	61	AGLARGEIAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDQLGADRWACMVGLARQPSV	120	
Db	61	AGLARGEIAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDQLGADRWACMVGLARQPSV	120	
QY	121	HPPLLVASFGTATTLDTIGPDNVPFGGLILPGPAMRGALAYGTAHLPLADGLVADYPID	180	
Db	121	HPPLLVASFGTATTLDTIGPDNVPFGGLILPGPAMRGALAYGTAHLPLADGLVADYPID	180	
QY	181	THQAIASGIAAAQAGAIVRQWLGRQRYGQAPEIYVAGGGWPEVRQEAERLLAVTGAAFG	240	
Db	181	THQAIASGIAAAQAGAIVRQWLGRQRYGQAPEIYVAGGGWPEVRQEAERLLAVTGAAFG	240	







QY 171 D--GL---VADYPIDTHQAIASGIAAAQAGAIVR 199  
| | :|:| | :|:| :|:| | | |  
Db 160 DPSGISPKIAEYGSSTEEAITAGVIHGLAGEIDR 193

RESULT 8  
US-09-813-453A-61  
; Sequence 61, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 61  
; LENGTH: 249  
; TYPE: PRT  
; ORGANISM: Pseudomonas putida  
US-09-813-453A-61

Query Match 11.7%; Score 163; DB 9; Length 249;  
Best Local Similarity 26.2%; Pred. No. 1e-06;  
Matches 76; Conservative 28; Mismatches 106; Indels 80; Gaps 11;  
QY 1 MIIIDSGNSRLKVGWFPDPAQOAREPAPVAFDNLDDLALGRWLATLPRPQALGVNV 60  
| | :|:| | | | :|:| :|:| | | |  
Db 1 MILELDCGNSFIK--WRVIHVADAVIEGGI-----VDS-----QALVAEV 40  
QY 61 AGLARGEIAIATLRAGGCDIRWL-----AQPLAM-----GLRNGY 96  
| | :|:| | | | :|:| :|:| | | |  
Db 41 AAL-----ASVRLTGCRIVSVRSEETDLCALIAQAFVQAKVAHPVREMAGVRNGY 93  
QY 97 RNPDLQADRWACMGV--LARQPSVHPPLLVASFATTTLDIGPDNVFPGLILPGPA 154  
: :|:| | | | :|:| | | | :|:| | | |  
Db 94 DDYQRLGMDRWLAALGAFHLAK-----GACLVIDLGTAAKADFVSADGEHLGGYICGMP 148  
QY 155 MMRGALAYGTAHLPLADGLVADYPI-----DTHQAIASGIAAAQAGAIVRQWLAGRQR 207  
:|:| | | :|:| :|:| :|:| | | |  
Db 149 LMRSQLRTHTRIRY-DDASAEALSSLSPOGRSTVEAVERGCVLMLQGFAYTQLEQARVL 207  
QY 208 YGOAPEIYVAGGWPQEVROEAERLLAVTGAAGFATPQPTYLDSPLVDGLA 257  
:|:| :|:| | | | :|:| :|:| | | |  
Db 208 WGEFTVFLTGADAPLVR-----AALPOARVVPDLVFFVGLA 243

RESULT 9  
US-09-813-453A-63  
; Sequence 63, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 63

; LENGTH: 241  
; TYPE: PRT  
; ORGANISM: Thiobacillus ferrooxidans  
US-09-813-453A-63  
Query Match 11.1%; Score 154.5; DB 9; Length 241;  
Best Local Similarity 28.6%; Pred. No. 5.6e-06;  
Matches 69; Conservative 27; Mismatches 108; Indels 37; Gaps 10;  
QY 2 IILIDSGNSRLKVG-----WFPDPAQOAREPAPVAFDNLDDLALG-RWLATLPRRQR 54  
:|:| | | | :|:| :|:| | | |  
Db 1 MIFIAVGNTRTLAHTHDGVHFDVSVVATSLPPTTEI-----LQOPGLTWLSAPNREPVA 54  
QY 55 ALGVNVAGLARGEIAIATLRAGGCDIRWLRAQPLAMGLRNGYRNPDLQADRWACMGVVL 114  
| | | | | | | :|:| :|:| | | |  
Db 55 LGGVVPAAALAAWREALATAEVREPDGFFR-----RAVPHDYHPPESLGFDRRCCLLAAA 109  
QY 115 ARQPSVHPPLLVASFATTTLDIGPDNVFPGLILPGPAMMRGALAYGTAHLPLADGLV 174  
| | :|:| | | | :|:| | | | | | |  
Db 110 MDYPG--QDSIVIDMGTAITIDLLAGGH-FRGGRIPLPGIAMSRLRGLHEGTALLP---EVV 163  
QY 175 ADYPI-----DTHQAIASGIAAAQAGAIVRQWLAGRQRYGOAPEIYVAGGWPQEVROEAE 229  
:|:| | | | :|:| :|:| | | |  
Db 164 LNAPAEMLGNDTSNAIQAGVIHLFADAL-RGAITDFRQYSPQARILITGG-----DAE 215  
QY 230 R 230  
|  
Db 216 R 216

RESULT 10  
US-09-813-453A-9  
; Sequence 9, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 246  
; TYPE: PRT  
; ORGANISM: Thermotoga maritima  
US-09-813-453A-9  
Query Match 10.9%; Score 152; DB 9; Length 246;  
Best Local Similarity 25.0%; Pred. No. 9.4e-06;  
Matches 62; Conservative 36; Mismatches 88; Indels 62; Gaps 10;  
QY 1 MIIIDSGNS-----RLKVGWFPDPAQOAREPAPVAFDNLDDLALGRW 44  
| | :|:| | | | :|:| :|:| | | |  
Db 1 MYLLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTE-----DELFSH 42  
QY 45 LATLPRRQALGVNVAGLARGEIAIATLRAGGCDI-----RWLRAQPLAMGLRNG---- 95  
| | | | | | | :|:| :|:| | | |  
Db 43 LHPLLGDAMREI-----KGIGVASVPTQNTVIERFSQKYFHISPIWVKAKNGCVKW 94  
QY 96 -YRNPDLQADRWACMGVVLARQPSVHPPLLVASFATTTLDIGPDNVFPGLILPGPA 154  
:|:| :|:| | | | :|:| | | | | | |  
Db 95 NVKNPSEVGADR---VANVAVFVKEYGKNGIIMGTATVLDLV-VNGSYEGGAILPGFF 150  
QY 155 MMRGALAYGTAHLPLADGLVADYPI--DTHQAIASGIAAAQAGAIVRQWLAGR--QRYGO 210  
| | :|:| | | | :|:| :|:| | | |  
Db 151 MMVHSLFRGTAKLPLVEVKPADFVVGKDTENIRLGVVNGSVYAL--EGIIIGRIKEVYGD 208



Db 161 LFARAANKLPRIEIAARPDDIIGKNTVSAMQAGILYGVGVQV--EGIVSRMKA KSKIPPVKVI 218

QY 216 VAGGGWPEVRQEAERL-----LAVTG 236

Db 219 ATGGLAPLIASESDIIDVVDPPFLTITG 245

RESULT 14

US-09-813-453A-59

; Sequence 59, Application US/09813453A

; Patent No. US20020168681A1

; GENERAL INFORMATION:

; APPLICANT: Yocum, R. Rogers

; APPLICANT: Patterson, Thomas A.

; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF

; TITLE OF INVENTION: ANTIBIOTICS

; FILE REFERENCE: OGZ-001

; CURRENT APPLICATION NUMBER: US/09/813,453A

; CURRENT FILING DATE: 2001-03-20

; PRIOR APPLICATION NUMBER: US 60/227,860

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: 09/667,569

; PRIOR FILING DATE: 2000-09-21

; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 59

; LENGTH: 212

; TYPE: PRT

; ORGANISM: Desulfovibrio vulgaris

US-09-813-453A-59

Query Match 9.9%; Score 138; DB 9; Length 212;

Best Local Similarity 27.4%; Pred. No. 0.00013;

Matches 62; Conservative 34; Mismatches 78; Indels 52; Gaps 12;

QY 3 ILIDSGNSRLKVGWFDPAQAPAPVAFDNLDLALGRWLATLPRRPQALGVNV-- 60

Db 6 LLEDIGNTNVKIG-----IAVETAVLTSY--VLPTDPGGQTTDSIGLRILLE 48

QY 61 ----AGLAR---GEAIAATLRAG-----GCDIRWLRAQ-----PLAMGLRNGYRNP 100

Db 49 VLRHAGLGPADVGVACVASSVPGVGNPLIRACE-RYLYRKLFLFAPGDIAIPLDNRYP 107

QY 101 QLGADRWACMVGVLARQPSVHPPLLLVA-SFGTATTLDTIGPDNVFPGLILPGPAMMRGA 159

Db 108 EVGADR--LVAAYAAARRLYPGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGA 164

QY 160 LAYGTAHLP-LADGLVADYPIDTHQAIAS-----GIAAAQAGAI 197

165 LSSRTAKLPRISLEVEEDSPVIGRSTTTSLNHGFI FGFAAMTEGVL 210

RESULT 15

US-09-813-453A-45

; Sequence 45, Application US/09813453A

; Patent No. US20020168681A1

; GENERAL INFORMATION:

; APPLICANT: Yocum, R. Rogers

; APPLICANT: Patterson, Thomas A.

; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF

; TITLE OF INVENTION: ANTIBIOTICS

; FILE REFERENCE: OGZ-001

; CURRENT APPLICATION NUMBER: US/09/813,453A

; CURRENT FILING DATE: 2001-03-20

; PRIOR APPLICATION NUMBER: US 60/227,860

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: 09/667,569

; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 45

; LENGTH: 262

; TYPE: PRT

; ORGANISM: Bacillus anthracis

US-09-813-453A-45

Query Match 9.9%; Score 138; DB 9; Length 262;

Best Local Similarity 23.7%; Pred. No. 0.00017;

Matches 62; Conservative 46; Mismatches 100; Indels 54; Gaps 11;

QY 1 MIILIDSGNSRLKVGWFDPAQAPAPVAFDNLDLALGRW-LATLPRRPQALGVN 59

Db 1 MIFVLDVGNNTNAVLGVFE-----EGELRQHWRMETDRHKTEDEYGML 42

QY 60 VAGLARGEAIAATLRAG-----GCDIRWLRAQPLAM--GLRNG----YR 97

Db 43 VKQLLEHEGLSFEDVKGIIVSSVVPPIMFALERMCE-KYFKIKPLVVGP GIKTGLNIKYE 101

QY 98 NPDQLGADRWACMVGVLARQPSVHPPLLLVASFGTATTLDTIGPDNVFPGLILPGPAMMR 157

Db 102 NPREVGADR---IVNAVAGIHLYGSPLIIVDFGTATTTCYINEEKHYMGGVITPGIMISA 158

QY 158 GALAYGTAHLPLADGLVADYPI--DTHQAIASGIA---AAQAGAIVRQWLAGRQRYGQAP 212

Db 159 EALYSRAAKLPRIETKPSVVGKNTVSAMQSGILYGVGVQVEGIVKRM---KEEAKQEP 215

QY 213 EIVVAGGGWPEVRQEAERLLAV 234

Db 216 KV-IATGGLAKLISEESNVIDV 236

Search completed: June 24, 2003, 22:24:03

Job time : 13.1733 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:26 ; Search time 27.4124 Seconds  
(without alignments)  
1297.879 Million cell updates/sec

Title: US-09-813-453A-15  
Perfect score: 1392  
Sequence: 1 MIILIDSGNSRLKVGWFDPD.....LDSVLDGLAALAAQGAPTA 267

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
arched: 908470 seqs, 133250620 residues 908470  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002:\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1392	100.0	267	AAU91162	Bordetella pertussis
2	258.5	18.6	455	AAU91162	Neisseria gonorrhoe
3	258.5	18.6	455	AAU91162	Neisseria gonorrhoe
4	258.5	18.6	460	AAU91162	Pantothenate kinas
5	258.5	18.6	592	AAU91162	Neisseria gonorrhoe
6	258.5	18.6	592	AAU91162	Neisseria gonorrhoe
7	256.5	18.4	455	AAU91162	Neisseria meningit
8	256.5	18.4	592	AAU91162	Neisseria meningit
9	256.5	18.4	592	AAU91162	Neisseria meningit
10	256.5	18.4	592	AAU91162	Pantothenate kinas

11	254.5	18.3	592	20	AAU91162	Neisseria meningit
12	254.5	18.3	592	21	AAU91162	Neisseria meningit
13	254.5	18.3	592	23	AAU91162	Pantothenate kinas
14	244	17.5	389	21	AAU91162	Neisseria meningit
15	194.5	14.0	242	23	AAU91162	Pantothenate kinas
16	186	13.4	189	20	AAU91162	Neisseria meningit
17	178.5	12.8	248	23	AAU91162	Pantothenate kinas
18	163	11.7	244	23	AAU91162	Pantothenate kinas
19	163	11.7	249	23	AAU91162	Listeria monocytog
20	161	11.6	259	23	AAU91162	Pantothenate kinas
21	154.5	11.1	241	23	AAU91162	Pantothenate kinas
22	152	10.9	246	23	AAU91162	Thermotoga maritim
23	148	10.6	255	23	AAU91162	Geobacter sulfurre
24	141	10.1	250	23	AAU91162	Clostridium acetob
25	140.5	10.1	258	23	AAU91162	Pantothenate kinas
26	138	9.9	212	23	AAU91162	Pantothenate kinas
27	138	9.9	262	23	AAU91162	Pantothenate kinas
28	136.5	9.8	257	23	AAU91162	Pantothenate kinas
29	134.5	9.7	260	23	AAU91162	Pantothenate kinas
30	134	9.6	258	23	AAU91162	Pantothenate kinas
31	134	9.6	265	23	AAU91162	Rhodobacter capsul
32	132	9.5	249	23	AAU91162	Streptomyces coeli
33	131.5	9.4	254	23	AAU91162	Pantothenate kinas
34	130.5	9.4	219	23	AAU91162	Pantothenate kinas
35	129.5	9.3	258	22	AAU91162	Pantothenate kinas
36	129.5	9.3	258	23	AAU91162	B. subtilis novel
37	127.5	9.2	256	23	AAU91162	Bacillus subtilis
38	125	9.0	233	23	AAU91162	Pantothenate kinas
39	121.5	8.7	548	22	AAU91162	Propionibacterium
40	112	8.0	9477	22	AAU91162	Streptomyces nous
41	105	7.5	1841	18	AAU91162	Tylactone synthase
42	105	7.5	4630	21	AAU91162	S. venezuelae vep
43	103.5	7.4	262	23	AAU91162	Deinococcus radiop
44	102.5	7.4	1996	18	AAU91162	Platenolide syntha
45	102.5	7.4	1996	18	AAU91162	Platenolide syntha

ALIGNMENTS

RESULT 1  
AAU91162  
ID AAU91162 standard; Protein; 267 AA.  
XX AAU91162;  
AC AAU91162;  
DT 05-JUN-2002 (first entry)  
XX Bordetella pertussis pantothenate kinase Coax.  
DE Pantothenate kinase; Coax; antibiotic; antimicrobial;  
XX pantothenate kinase modulator; coenzyme A; bactericidal compound.  
KW Bordetella pertussis.  
OS WO200216601-A2.  
XX 28-FEB-2002.  
XX 24-AUG-2001; 2001WO-US26531.  
XX 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX Yocum RE, Patterson TA;  
XX WPI; 2002-269358/31.  
XX Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -

```
XX PS Claim 6; Page 79-80; 128pp; English.
XX CC The invention describes assays for identifying a (potential) antibiotic
XX CC comprising contacting an assay composition comprising a pantothenate
XX CC kinase (Coax) protein with a test compound, and determining the ability
XX CC of the test compound to inhibit the activity of the Coax protein, an
XX CC essential enzyme for the production of coenzyme A. Coax protein is a
XX CC valuable target for identifying bactericidal compounds. Coax modulating
XX CC agents can be used in an infectious animal model to determine the
XX CC efficacy, toxicity, or side effects of treatment with such an agent. This
XX CC is the amino acid sequence of a pantothenate kinase (Coax) protein
XX CC described in the invention.
XX SQ Sequence 267 AA;
Query Match 100.0%; Score 1392; DB 23; Length 267;
Best Local Similarity 100.0%; Pred. No. 7.2e-139;
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 MIILIDSGNSRLKVGWFDPAQAPAPVAFDNLDLALGRWLATLPRRPQALGVNV 60
1 MIILIDSGNSRLKVGWFDPAQAPAPVAFDNLDLALGRWLATLPRRPQALGVNV 60
61 AGLARGEIAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDLGADRWACMVGLARQPSV 120
61 AGLARGEIAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDLGADRWACMVGLARQPSV 120
121 HPPLLVSFGTATTLDITIGPDNVFPGGLILPGPMMRGALAYGTAHLPLADGLVADYPI 180
121 HPPLLVSFGTATTLDITIGPDNVFPGGLILPGPMMRGALAYGTAHLPLADGLVADYPI 180
181 THQAIASGIAAAQAGAIYRQWLAGRQRYGQAPAIYVAGGWPVEVQEAERLLAVTGA 240
181 THQAIASGIAAAQAGAIYRQWLAGRQRYGQAPAIYVAGGWPVEVQEAERLLAVTGA 240
241 ATPQPTYLDSPLVDGLAALAAQCAPTA 267
241 ATPQPTYLDSPLVDGLAALAAQCAPTA 267
```

```
RESULT 2
AAY38617
ID AAY38617 standard; Protein; 455 AA.
XX AC AAY38617;
XX DT 08-OCT-1999 (first entry)
XX Neisseria gonorrhoeae antigen encoded by partial ORF61.
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX OS Neisseria gonorrhoeae.
XX WO9924578-A2.
XX PD 20-MAY-1999.
XX PF 09-OCT-1998; 98WO-IB01665.
XX PR 01-SEP-1998; 98GB-0019016.
XX PR 06-NOV-1997; 97GB-0023516.
XX PR 14-NOV-1997; 97GB-0024190.
XX PR 18-NOV-1997; 97GB-0024386.
XX PR 27-NOV-1997; 97GB-0025158.
XX PR 10-DEC-1997; 97GB-0026147.
XX PR 14-JAN-1998; 98GB-0000759.
XX (CHIR-) CHIRON SPA.
XX PA Grandi G, Massignani V, Pizza M, Rappuoli R, Scarlato V;
XX PI
```

```
XX WPI; 1999-327407/27.
XX PT Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
XX PT diagnosis, treatment and prevention of infection
XX PS Claim 4; Page 175; 524pp; English.
XX CC Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
XX CC and N. gonorrhoeae antigenic proteins. They are encoded by open
XX CC reading frames (ORFs) AAZ11972-212358. The antigenic proteins,
XX CC their fragments, their nucleic acids and antibodies are used for
XX CC diagnosis, prevention (as vaccines) or treatment of Neisseria
XX CC infections, such as meningitis, septicaemia and gonorrhea. Both
XX CC organisms are closely related. Fragments of the nucleic acids
XX CC are useful as hybridisation probes and antisense reagents.
XX SQ Sequence 455 AA;
Query Match 18.6%; Score 258.5; DB 20; Length 455;
Best Local Similarity 31.8%; Pred. No. 1.6e-18;
Matches 84; Conservative 43; Mismatches 116; Indels 21; Gaps 8;
3 ILIDSGNSRLKVGWFDPAQAPAPVAFDNLDLALGRWLATLPRRPQALGVNVAG 62
204 LLEGGNSRLKWAWE-NGTFATVGSAPYR---DLSPGAEWAKEADGNVRIVGCAVCG 258
63 LARGEIAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDLGADRWACMVGLARQPSVHP 122
259 ESKKAQVKEQLAR---KIEWLPSSAQALGIRNHYRHPHEHSDRWFNALG--SRRFS-RN 312
123 PLLVASFGTATTLDITIGPDNVFPGGLILPGPMMRGALAYGTAHLPLADGLVADYPI 182
313 ACVVVSCGTAVTVDALTDGHLGGTIMPGFHLMKESLAVRTANLNRPAKRYPPFTTTG 372
183 QAIASGIAAAQAGAIYRQW--LAGRQRYGQAPAIYVAGGWPVEVQEAERLLAVTGA 240
373 NAVASGMDAVCGSIMMHGRLKEKNGAGKPDVVIITGGG-----AAKVAEALPPAFL 425
241 ATPQPTYLDSPLVDGLAAL-AAQG 263
426 AENTVRVADNLVIHGLLNLI AEG 449
RESULT 3
AAY74908
ID AAY74908 standard; Protein; 455 AA.
XX AC AAY74908;
XX DT 21-MAR-2000 (first entry)
XX Neisseria gonorrhoeae ORF 311 protein sequence SEQ ID NO:1290.
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy.
XX OS Neisseria gonorrhoeae.
XX PN WO9957280-A2.
XX PD 11-NOV-1999.
XX PF 30-APR-1999; 99WO-US09346.
XX PR 01-MAY-1998; 98US-0083758.
XX PR 31-JUL-1998; 98US-0094869.
XX PR 02-SEP-1998; 98US-0098994.
XX PR 02-SEP-1998; 98US-0099062.
XX PR 09-OCT-1998; 98US-0103749.
XX PR 09-OCT-1998; 98US-0103794.
XX PR 09-OCT-1998; 98US-0103796.
```



```
OS Neisseria gonorrhoeae.
XX WO9924578-A2.
XX 20-MAY-1999.
XX 09-OCT-1998; 98WO-IB01665.
XX 01-SEP-1998; 98GB-0019016.
PR 06-NOV-1997; 97GB-0023516.
PR 14-NOV-1997; 97GB-0024190.
PR 18-NOV-1997; 97GB-0024386.
PR 27-NOV-1997; 97GB-0025158.
PR 10-DEC-1997; 97GB-0026147.
PR 14-JAN-1998; 98GB-0000759.
XX (CHIR-) CHIRON SPA.
XX Grandi G, Massignani V, Pizza M, Rappuoli R, Scarlato V;
PI WPI; 1999-327407/27.
XX N-PSDB; AAZ12075.
XX Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT diagnosis, treatment and prevention of infection
XX Claim 4; Page 175; 524pp; English.
XX Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC and N. gonorrhoeae antigenic proteins. They are encoded by open
CC reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC their fragments, their nucleic acids and antibodies are used for
CC diagnosis, prevention (as vaccines) or treatment of Neisseria
CC infections, such as meningitis, septicaemia and gonorrhea. Both
CC organisms are closely related. Fragments of the nucleic acids
CC are useful as hybridisation probes and antisense reagents.
XX
SQ Sequence 592 AA;
Query Match 18.6%; Score 258.5; DB 20; Length 592;
Best Local Similarity 31.8%; Pred. No. 2.3e-18;
Matches 84; Conservative 43; Mismatches 116; Indels 21; Gaps 8;
QY 3 ILIDSGNSRLKVGWFDPAQOAREPAPVAFDNLDDLGRWLATLPRRQALGVNVAG 62
Db 341 LLEGGNSRLKAWVE-NGTFATVGSAPYR---DLSPGLAEWAEEKADGNVRIVGCAVCG 395
QY 63 LARGEIAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDLGADRWACMVGVLARQPSVHP 122
~ 396 ESKKAQVKEQLAR---KIEWLPSSAQALGIRNHYRHPHEHGSDFWENALG--SRRFS-RN 449
QY 123 PLLVASFGTATLDTIGPDNVFPFGLILPGPAMRGALAYGTAHLPLADGLVADYPIDTH 182
Db 450 ACVVVSCGTAVTVDALTDGHLGGTIMGPFHLMKESLAVRTANLNRPAKRYPFPTTTG 509
QY 183 QAIASGIAAQAQAIVRQW--LAGRQRYGQAPFIYVAGGGWPEVROEAERLLAVTGAAG 240
Db 510 NAVASGMMDAVCGSIMMHGRLKEKNGACKPVDVIITGGG-----AAKVAEALPPAFL 562
QY 241 ATPQPTVLDSPVLDGLAAL-AAQG 263
Db 563 AENTVRVADNLVIHGLLNLIAAEG 586
RESULT 6
AAY74911
ID AAY74911 standard; Protein; 592 AA.
XX
AC AAY74911;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria gonorrhoeae ORF 311 protein sequence SEQ ID NO:1296.
```

```
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy.
XX
OS Neisseria gonorrhoeae.
XX
PN WO9957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
DR WPI; 2000-062150/05.
DR N-PSDB; AAZ53673.
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
XX vaccines and diagnostics
XX
PS Claim 2; Page 713; 1453pp; English.
XX
CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 592 AA;
Query Match 18.6%; Score 258.5; DB 21; Length 592;
Best Local Similarity 31.8%; Pred. No. 2.3e-18;
Matches 84; Conservative 43; Mismatches 116; Indels 21; Gaps 8;
QY 3 ILIDSGNSRLKVGWFDPAQOAREPAPVAFDNLDDLGRWLATLPRRQALGVNVAG 62
Db 341 LLEGGNSRLKAWVE-NGTFATVGSAPYR---DLSPGLAEWAEEKADGNVRIVGCAVCG 395
QY 63 LARGEIAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDLGADRWACMVGVLARQPSVHP 122
Db 396 ESKKAQVKEQLAR---KIEWLPSSAQALGIRNHYRHPHEHGSDFWENALG--SRRFS-RN 449
QY 123 PLLVASFGTATLDTIGPDNVFPFGLILPGPAMRGALAYGTAHLPLADGLVADYPIDTH 182
Db 450 ACVVVSCGTAVTVDALTDGHLGGTIMGPFHLMKESLAVRTANLNRPAKRYPFPTTTG 509
QY 183 QAIASGIAAQAQAIVRQW--LAGRQRYGQAPFIYVAGGGWPEVROEAERLLAVTGAAG 240
Db 510 NAVASGMMDAVCGSIMMHGRLKEKNGACKPVDVIITGGG-----AAKVAEALPPAFL 562
```







CC The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.

XX  
SQ Sequence 592 AA;  
Query Match 18.4%; Score 256.5; DB 23; Length 592;  
Best Local Similarity 32.0%; Pred. No. 3.7e-18;  
Matches 85; Conservative 43; Mismatches 113; Indels 25; Gaps 9;

QY 3 ILIDSGNSRLKVGWFDPAQAAAREPAPVAFDNLDDLALGRWLATLPRRPQALGVNVAG 62  
:|:| | | | | | : : : | | | | | | | | :|:| | |  
341 LLLDGGNSRLKWAWE-NGTFATVGSAPYR----DLSPLGAWEAEKVDGNVIRIVGCAVCG 395  
QY 63 LARGEIAIATLRAGGCDIRWLRAQPLAMGLRNGYRNPDLGADRWACMVGVLARQPSVHP 122  
: : : | | | : | | | | | | | | :|:| | :|:| |  
396 EFKAQVQEQLAR---KIEWLPSSAQALGIRNHYRHPPEHGSDFWFNALG--SRRFS-RN 449  
QY 123 PLLVASEFGTATLDTIGPDNVFPFGLILPGPAMMRGALAYGTAHLPLADGLVADYPIDTH 182  
:|:| | | | | : : : | | | | | | | | :|:| | :|:| |  
450 ACVVVSCGTAVTVDAITDGHYLGTTMPGFHLMKESLAVRTANLNHRHAGKRYPFPTTTG 509  
QY 183 QAIASGIAAAQAGAIVRQWLAGRQR----YQQAPEIYVAGGGWPEVQEAERLLAVTGAA 238  
:|:| | :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|  
510 NAVASGMMDAVCGSVM--MMHGRLEKKTGAGKPVVDVIITGGG-----AAKVAEALPPA 560  
QY 239 FGATPQPTYLDSPVLDGLAAL-AAQG 263  
:|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|  
561 FLAENTVRVADNLVIHGLLNLIAAEG 586

RESULT 11  
AAY38615  
ID AAY38615 standard; Protein; 592 AA.  
XX  
AC AAY38615;  
XX  
DT 08-OCT-1999 (first entry)  
XX  
DE Neisseria meningitidis antigen encoded by ORF61.  
W Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.  
XX  
OS Neisseria meningitidis.  
XX  
PN WO9924578-A2.  
XX  
PD 20-MAY-1999.  
XX  
PF 09-OCT-1998; 98WO-IB01665.  
XX  
PR 01-SEP-1998; 98GB-0019016.  
PR 06-NOV-1997; 97GB-0023516.  
PR 14-NOV-1997; 97GB-0024190.  
PR 18-NOV-1997; 97GB-0024386.  
PR 27-NOV-1997; 97GB-0025158.  
PR 10-DEC-1997; 97GB-0026147.  
PR 14-JAN-1998; 98GB-0000759.  
XX  
PA (CHIR-) CHIRON SPA.

XX  
PI Grandi G, Massignani V, Pizza M, Rappuoli R, Scarlato V;  
XX  
DR WPI; 1999-327407/27.  
DR N-PSDB; AAZ12073.

XX  
PT Proteins from Neisseria meningitidis and N. gonorrhoeae useful for  
PT diagnosis; treatment and prevention of infection  
XX  
PS Claim 4; Page 172; 524pp; English.  
XX  
CC Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis  
CC and N. gonorrhoeae antigenic proteins. They are encoded by open  
CC reading frames (ORFs) AAZ11972-212358. The antigenic proteins,  
CC their fragments, their nucleic acids and antibodies are used for  
CC diagnosis, prevention (as vaccines) or treatment of Neisseria  
CC infections, such as meningitis, septicaemia and gonorrhea. Both  
CC organisms are closely related. Fragments of the nucleic acids  
CC are useful as hybridisation probes and antisense reagents.  
XX  
SQ Sequence 592 AA;

Query Match 18.3%; Score 254.5; DB 20; Length 592;  
Best Local Similarity 31.6%; Pred. No. 6e-18;  
Matches 84; Conservative 44; Mismatches 113; Indels 25; Gaps 9;  
QY 3 ILIDSGNSRLKVGWFDPAQAAAREPAPVAFDNLDDLALGRWLATLPRRPQALGVNVAG 62  
:|:| | | | | | : : : | | | | | | | | :|:| | |  
341 LLLDGGNSRLKWAWE-NGTFATVGSAPYR----DLSPLGAWEAEKADGNVIRIVGCAVCG 395  
QY 63 LARGEIAIATLRAGGCDIRWLRAQPLAMGLRNGYRNPDLGADRWACMVGVLARQPSVHP 122  
: : : | | | : | | | | | | | | :|:| | :|:| |  
396 EFKAQVQEQLAR---KIEWLPSSAQALGIRNHYRHPPEHGSDFWFNALG--SRRFS-RN 449  
QY 123 PLLVASEFGTATLDTIGPDNVFPFGLILPGPAMMRGALAYGTAHLPLADGLVADYPIDTH 182  
:|:| | | | | : : : | | | | | | | | :|:| | :|:| |  
450 ACVVVSCGTAVTVDAITDGHYLGTTMPGFHLMKESLAVRTANLNHRHAGKRYPFPTTTG 509  
QY 183 QAIASGIAAAQAGAIVRQWLAGRQR----YQQAPEIYVAGGGWPEVQEAERLLAVTGAA 238  
:|:| | :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|  
510 NAVASGMMDAVCGSVM--MMHGRLEKKTGAGKPVVDVIITGGG-----AAKVAEALPPA 560  
QY 239 FGATPQPTYLDSPVLDGLA-ALAAQG 263  
:|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|  
561 FLAENTVRVADNLVIYGLLNMIAAEG 586

RESULT 12  
AAY74912  
ID AAY74912 standard; Protein; 592 AA.  
XX  
AC AAY74912;  
XX  
DT 21-MAR-2000 (first entry)  
XX  
DE Neisseria meningitidis ORF 311 protein sequence SEQ ID NO:1298.  
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;  
KW antibacterial; gene therapy.  
XX  
OS Neisseria meningitidis.  
XX  
PN WO9957280-A2.  
XX  
PD 11-NOV-1999.  
XX  
PF 30-APR-1999; 99WO-US09346.  
XX  
PR 01-MAY-1998; 98US-0083758.  
PR 31-JUL-1998; 98US-0094869.  
PR 02-SEP-1998; 98US-0098994.  
PR 02-SEP-1998; 98US-0099062.  
PR 09-OCT-1998; 98US-0103749.  
PR 09-OCT-1998; 98US-0103794.  
PR 09-OCT-1998; 98US-0103796.  
PR 25-FEB-1999; 99US-0121528.  
XX

PA (CHIR ) CHIRON CORP.  
PA (GENO-) INST GENOMIC RES.  
XX  
PI Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;  
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
PI Tettelin H, Venter JC;  
XX  
DR WPI; 2000-062150/05.  
DR N-PSDB; AAZ53674.  
XX  
PT Novel Neisserial polypeptides predicted to be useful antigens for  
PT vaccines and diagnostics  
XX  
PS Claim 2; Page 714; 1453pp; English.  
XX  
CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941  
CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides  
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent  
CC PCR primers used in the exemplification of the present inventions. The  
CC polypeptides, the polynucleotides, antibodies and compositions of  
CC the invention can be used as vaccines, as diagnostic reagents, and as  
CC immunogenic compositions. The polypeptides can be used in the  
CC manufacture of medicaments for treating or preventing infection due to  
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the  
CC presence of Neisseria bacteria, or to raise antibodies. They may also  
CC be used to screen for agonists or antagonists, which may themselves  
CC have use as antibacterial agents. The polynucleotides of the invention  
CC may also be used in gene therapy protocols.  
XX  
SQ Sequence 592 AA;

Query Match 18.3%; Score 254.5; DB 21; Length 592;  
Best Local Similarity 31.6%; Pred. No. 6e-18;  
Matches 84; Conservative 44; Mismatches 113; Indels 25; Gaps 9;

QY 3 ILIDSGNSRLKVGWFDPAQAPAPVAFDNLDDLGRWLATLPRRQALGVNVAG 62  
Db 341 LLLDGGNSRLKWAWE-NGTFATVGSAPYR----DLSPLGAEWAEEKADGNVRIVGCAVCG 395  
QY 63 LARGEIAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDLGADRWACMVGLARQPSVHP 122  
Db 396 EFKAQVQEQLAR---KIEWLPSSAQALGIRNHYRHPPEHGSDFWENALG--SRRES-RN 449  
QY 123 PLLVASFGTATTLDITGPDNVFPGGLILPGPAMRGALAYGTAHLPLADGLVADYPIDTH 182  
Db 450 ACVVVSCGTAVTVDALTDGHLGGTIMPGFHLMKESLAVRTANLNRHAGKRYPPFTTTG 509  
QY 183 QAIASGIAAAQAGAIVRQWLAGRQR----YGOAPEIYVAGGWPVEVQEAERLLAVTGAA 238  
Db 510 NAVASGMMDAVCGSVN--MMHGRLEKKTGAGKPDVDIITGGG-----AAKVAEALPPA 560  
QY 239 FGATPQPTYLDSPVLDGLA-ALAAQG 263  
Db 561 FLAENTVRVADNLVIYGLLNMIAAEG 586

RESULT 13  
AAU91169  
ID AAU91169 standard; Protein; 592 AA.  
XX  
AC AAU91169;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Pantothenate kinase (Coax) #7.  
XX  
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX  
OS Neisseria meningitidis.  
XX  
PN WO200216601-A2.  
XX

PD 28-FEB-2002.  
XX  
PF 24-AUG-2001; 2001WO-US26531.  
XX  
PR 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX  
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX  
PI Yocum RR, Patterson TA;  
XX  
DR WPI; 2002-269358/31.  
DR N-PSDB; ABK54190.  
XX  
PT Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein  
XX  
PS Claim 6; Page 96-98; 128pp; English.  
XX  
CC The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.  
XX  
SQ Sequence 592 AA;

Query Match 18.3%; Score 254.5; DB 23; Length 592;  
Best Local Similarity 31.6%; Pred. No. 6e-18;  
Matches 84; Conservative 44; Mismatches 113; Indels 25; Gaps 9;

QY 3 ILIDSGNSRLKVGWFDPAQAPAPVAFDNLDDLGRWLATLPRRQALGVNVAG 62  
Db 341 LLLDGGNSRLKWAWE-NGTFATVGSAPYR----DLSPLGAEWAEEKADGNVRIVGCAVCG 395  
QY 63 LARGEIAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDLGADRWACMVGLARQPSVHP 122  
Db 396 EFKAQVQEQLAR---KIEWLPSSAQALGIRNHYRHPPEHGSDFWENALG--SRRES-RN 449  
QY 123 PLLVASFGTATTLDITGPDNVFPGGLILPGPAMRGALAYGTAHLPLADGLVADYPIDTH 182  
Db 450 ACVVVSCGTAVTVDALTDGHLGGTIMPGFHLMKESLAVRTANLNRHAGKRYPPFTTTG 509  
QY 183 QAIASGIAAAQAGAIVRQWLAGRQR----YGOAPEIYVAGGWPVEVQEAERLLAVTGAA 238  
Db 510 NAVASGMMDAVCGSVN--MMHGRLEKKTGAGKPDVDIITGGG-----AAKVAEALPPA 560  
QY 239 FGATPQPTYLDSPVLDGLA-ALAAQG 263  
Db 561 FLAENTVRVADNLVIYGLLNMIAAEG 586

RESULT 14  
AAY74909  
ID AAY74909 standard; Protein; 389 AA.  
XX  
AC AAY74909;  
XX  
DT 21-MAR-2000 (first entry)  
XX  
DE Neisseria meningitidis ORF 311 protein sequence SEQ ID NO:1292.  
XX  
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;  
KW antibacterial; gene therapy.  
XX  
OS Neisseria meningitidis.



```
XX WO9957280-A2.
PN 11-NOV-1999.
XX 30-APR-1999; 99WO-US09346.
XX 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
XX Tettelin H, Venter JC;
DR WPI; 2000-062150/05.
DR N-PSDB; AAZ53671.
XX Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics
XX Claim 2; Page 710; 1453pp; English.
XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC represent novel Neisseria meningitis and N. gonorrhoea polynucleotides
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX Sequence 389 AA;
SQ
Query Match 17.5%; Score 244; DB 21; Length 389;
Best Local Similarity 31.5%; Pred. No. 4.3e-17;
Matches 84; Conservative 44; Mismatches 113; Indels 26; Gaps 10;
QY 3 ILIDSGNSRLKVGWFDPAQAAREPAPVAFDNLDLALGRWLATLPRRPQALGVNVAG 62
Db 138 LLLDGGNSRLKWAWE-NGTFATVGSAPYR---DLSPGLAEWAEEKADGNVRIVGCACVG 192
QY 63 LARGEAIAATLRAGGCDINWLRQAQLAM-GLRNGYRNPDLQADRWACMVGLARQPSVH 121
Db 193 EFKAQVQEQALR---KIEWLPSSAQALFGIRNHYRHPHEHSGSDRWFNALG--SRRES-R 246
QY 122 PPLLVSFGTATTLDTIGPDNVFPGGILPGPAMMRGALAYGTAHLPLADGLVADYDIT 181
Db 247 NACVVVSCGTAVTVDDTDGHYLGTTIMPGFHLMKESLAVRTANLNRHAGKRYPFPTT 306
QY 182 HQAIASGIAAAQAGAIYRQWLAGRQR---YGOAPEIIVAGGGWPEVQEAERLLAVTGA 237
Db 307 GNAVASGMDAVCGSV--MMHGRLEKKTGAGKPVVDVITGGG-----AAKVAEALPP 357
QY 238 AFGATPQPTYLDSPVLDGLA-ALAAQG 263
Db 358 AFLAENTVRVADNLVIYGLNMIAAEG 384
```

RESULT 15

```
AAU91180
ID AAU91180 standard; Protein; 242 AA.
XX
AC AAU91180;
XX 05-JUN-2002 (first entry)
XX Pantothenate kinase (Coax) #18.
DE Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX Xylessa fastidiosa.
OS WO200216601-A2.
XX 28-FEB-2002.
XX 24-AUG-2001; 2001WO-US26531.
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX Yocum RR, Patterson TA;
PI WPI; 2002-269358/31.
XX N-PSDB; ABK54201.
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein -
XX Claim 6; Page 111; 128pp; English.
XX The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX
SQ Sequence 242 AA;
Query Match 14.0%; Score 194.5; DB 23; Length 242;
Best Local Similarity 27.9%; Pred. No. 3.9e-12;
Matches 80; Conservative 25; Mismatches 105; Indels 77; Gaps 10;
QY 4 LIDSGNSRLKVGWFDPAQAAREPA-----PVAFDNLDLALGRWLATLPRRPQAL 56
Db 5 LFDLGNSEFKCA-----SLREGVIGPVTVLPYLTETMDAFA----LQELPRG----- 47
QY 57 GVNVALARGEA-----IAATLRAGGCDIRWLRQAQLAMGLRNGYR 97
Db 48 --RVAYLASVAAPAITHVLEVLKIHFEEQVQAATVAA-----CAGVRIAYA 92
QY 98 NPDQLGADRWACMVGLARQPSVHPPLLVASFSGTATTLDTIGPDNVFPGGILPGPAMMR 157
Db 93 HPERFGVDRFLALLGSYG-----EGNVLVVGVGTALTIDLLAANGCHLGGRIASPTLMR 147
QY 158 GALAYGTAHLPLADGLVADYPIDTHQATASGIAAAQAGAIYRQWLAGRQRYGQAPEIYVA 217
Db 148 QALHARAEQLPLSGGNYLEFAEDTEDALVSGCGAAGAAVALIERSLYEAHORLDQSVRLLLH 207
QY 218 GGGWPEVQEAERLLAVTGAAGFATPQPTFYLDSPVLDGLAALAAQGA 264
Db 203 GGG-----VASLLPWLG---DVVHRPTL-----VLDGLAIWAATAA 240
```

Search completed: June 24, 2003, 21:46:23  
Job time : 29.4124 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:36:04 ; Search time 9.7386 Seconds  
(without alignments)  
2635.685 Million cell updates/sec

Title: US-09-813-453A-15  
Perfect score: 1392  
Sequence: 1 MIILIDSGNSRLKVGWFDPD.....LDSPVLDGLAALAAQGAPTA 267

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1392	100.0	267	2 I40327	baf protein - Bord
2	256.5	18.4	592	2 H82031	probable biotin-[a
3	254.5	18.3	592	2 B81009	BirA protein/Bvg a
4	194.5	14.0	242	2 A82637	conserved hypothet
5	178.5	12.8	248	2 H83111	hypothetical prote
6	161	11.6	259	2 AF1102	conserved hypothet
7	160.5	11.5	273	2 E97293	probable transcrip
8	155	11.1	259	2 AF1464	conserved hypothet
9	152	10.9	246	2 D72320	conserved hypothet
10	134.5	9.7	261	2 B87489	transcription acti
11	134	9.6	265	2 T36391	hypothetical prote
12	131.5	9.4	254	2 F83660	hypothetical prote
13	125	9.0	233	2 S66100	conserved hypothet
14	122	8.8	251	2 F83604	hypothetical prote
15	110.5	7.9	591	2 T44868	probable membrane
16	110.5	7.9	600	2 D87232	conserved membrane
17	109	7.8	636	2 T18542	mofC protein - Lep
18	107.5	7.7	480	2 F70785	hypothetical prote
19	107	7.7	4848	2 T30289	pristinamycin I sy
20	106.5	7.7	423	2 A36794	hypothetical prote
21	103.5	7.4	262	2 E75516	conserved hypothet
22	103	7.4	974	2 T35045	bacteriophage phiC
23	102.5	7.4	442	1 XUBRVS	3-phosphoshikimate
24	102.5	7.4	465	2 H87301	conserved hypothet
25	102.5	7.4	603	2 A87342	conserved hypothet
26	101	7.3	309	2 B75565	hypothetical prote
27	101	7.3	566	2 T35203	probable two-compo
28	101	7.3	1110	1 S55279	guanylate cyclase
29	100.5	7.2	546	2 D83408	hypothetical prote

30	100.5	7.2	813	2 E87396	helicase, probable
31	100.5	7.2	821	2 AB0688	probable bacteriop
32	100	7.2	308	2 F95903	conserved hypothet
33	99.5	7.1	240	2 G84177	hypothetical prote
34	99.5	7.1	274	2 H86937	conserved hypothet
35	99.5	7.1	700	2 A84243	phosphoribosylform
36	99	7.1	280	2 B84219	4-hydroxybenzoate
37	99	7.1	509	2 T34871	probable membrane
38	99	7.1	974	2 T30204	pgl2 protein - Str
39	98.5	7.1	676	1 EDBE22	immediate-early pr
40	98	7.0	439	2 I57561	transcription fact
41	98	7.0	509	2 E85840	hypothetical prote
42	97.5	7.0	7576	2 T17428	FK506 polyketide s
43	97	7.0	257	2 S75559	hypothetical prote
44	97	7.0	509	2 H90994	hypothetical prote
45	97	7.0	562	2 H70688	hypothetical prote

ALIGNMENTS

RESULT 1

I40327

baf protein - Bordetella pertussis

C;Species: Bordetella pertussis

C;Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 08-Oct-1999

C;Accession: I40327; S70669

R;DeShazer, D.; Wood, G.E.; Friedman, R.L.

J. Bacteriol. 177, 3801-3807, 1995

A;Title: Identification of a Bordetella pertussis regulatory factor required for tr

A;Reference number: I40327; MUID:95325323; PMID:7601846

A;Accession: I40327

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-267 <RES>

A;Cross-references: EMBL:U12020; NID:g687228; PIDN:AAA75361.1; PID:g687229

R;Allen, A.; Maskell, D.

Mol. Microbiol. 19, 37-52, 1996

A;Title: The identification, cloning and mutagenesis of a genetic locus required for

A;Reference number: S70669; MUID:96419162; PMID:8821935

A;Accession: S70669

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 239-267 <ALL>

A;Cross-references: EMBL:X90711; NID:g992967; PIDN:CAA62242.1; PID:g992968

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

C;Genetics:

A;Gene: baf

Query Match 100.0%; Score 1392; DB 2; Length 267;  
Best Local Similarity 100.0%; Pred. No. 7.2e-103;  
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MIILIDSGNSRLKVGWFDPAQAAREPAPVAFDNLDLALGRWLATLPRRPQALGVNV	60
Db	1	MIILIDSGNSRLKVGWFDPAQAAREPAPVAFDNLDLALGRWLATLPRRPQALGVNV	60
QY	61	AGLARGEIAAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDLGADRWACMVGLARQPSV	120
Db	61	AGLARGEIAAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDLGADRWACMVGLARQPSV	120
QY	121	HPPLLVASFGTATTLDITIGPDNVFPGGLILPGPAMRGALAYGTAHLPLADGLVADYPID	180
Db	121	HPPLLVASFGTATTLDITIGPDNVFPGGLILPGPAMRGALAYGTAHLPLADGLVADYPID	180
QY	181	THQAIASGIAAAQAGAIVRQWLGRQRYGQAPFIYVAGGGWPEVRQEAERLLAVTGAAGF	240
Db	181	THQAIASGIAAAQAGAIVRQWLGRQRYGQAPFIYVAGGGWPEVRQEAERLLAVTGAAGF	240
QY	241	ATPQPTYLDSPVLDGLAALAAQGAPTA	267
Db	241	ATPQPTYLDSPVLDGLAALAAQGAPTA	267















GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:08:46 ; Search time 10.4407 Seconds  
(without alignments)  
2373.336 Million cell updates/sec

Title: US-09-813-453A-12  
Perfect score: 1168  
Sequence: 1 MRFLTVGVNSSVDIALWEG.....YDPLLVHGRNLLLYHRI 229

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

---earched: 417779 seqs, 108206813 residues 417779  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :				Published_Applications_AA:*	
1:	/cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*				
2:	/cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*				
3:	/cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*				
4:	/cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*				
5:	/cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*				
6:	/cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*				
7:	/cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*				
8:	/cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*				
9:	/cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*				
10:	/cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*				
11:	/cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*				
12:	/cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*				
13:	/cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*				
14:	/cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*				

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1168	100.0	229	9 US-09-813-453A-12	Sequence 12, Appl
2	225	19.3	257	9 US-09-813-453A-53	Sequence 53, Appl
3	193.5	16.6	246	9 US-09-813-453A-9	Sequence 9, Appl
4	177.5	15.2	209	9 US-09-813-453A-21	Sequence 21, Appl
5	167.5	14.3	262	9 US-09-813-453A-45	Sequence 45, Appl
6	164.5	14.1	241	9 US-09-813-453A-63	Sequence 63, Appl
7	158.5	13.6	258	9 US-09-813-453A-6	Sequence 6, Appl
8	154.5	13.2	258	9 US-09-813-453A-2	Sequence 2, Appl
9	148.5	12.7	258	9 US-09-813-453A-49	Sequence 49, Appl
10	141.5	12.1	262	9 US-09-813-453A-11	Sequence 11, Appl
11	139.5	11.9	255	9 US-09-813-453A-7	Sequence 7, Appl
12	138.5	11.9	223	9 US-09-895-913A-74	Sequence 74, Appl
13	138.5	11.9	223	9 US-09-813-453A-14	Sequence 14, Appl
14	138.5	11.9	223	9 US-09-813-453A-67	Sequence 67, Appl
15	138.5	11.9	254	9 US-09-813-453A-47	Sequence 47, Appl
16	133.5	11.4	262	9 US-09-813-453A-8	Sequence 8, Appl
17	133	11.4	256	9 US-09-813-453A-55	Sequence 55, Appl
18	133	11.4	257	9 US-09-813-453A-13	Sequence 13, Appl
19	132.5	11.3	244	9 US-09-813-453A-41	Sequence 41, Appl

20	130	11.1	219	9 US-09-813-453A-57	Sequence 57, Appl
21	128.5	11.0	265	9 US-09-813-453A-4	Sequence 4, Appl
22	125.5	10.7	260	9 US-09-813-453A-51	Sequence 51, Appl
23	123.5	10.6	250	9 US-09-813-453A-3	Sequence 3, Appl
24	122.5	10.5	273	9 US-09-813-453A-10	Sequence 10, Appl
25	120.5	10.3	212	9 US-09-813-453A-59	Sequence 59, Appl
26	117	10.0	233	9 US-09-813-453A-17	Sequence 17, Appl
27	114	9.8	272	9 US-09-712-363-276	Sequence 276, App
28	114	9.8	272	9 US-09-813-453A-5	Sequence 5, Appl
29	108	9.2	592	9 US-09-813-453A-22	Sequence 22, Appl
30	103	8.8	592	9 US-09-813-453A-43	Sequence 43, Appl
31	99	8.5	460	9 US-09-813-453A-39	Sequence 39, Appl
32	98.5	8.4	248	9 US-09-813-453A-20	Sequence 20, Appl
33	95	8.1	242	9 US-09-813-453A-65	Sequence 65, Appl
34	92.5	7.9	249	9 US-09-813-453A-61	Sequence 61, Appl
35	89.5	7.7	249	9 US-09-813-453A-70	Sequence 70, Appl
36	87.5	7.5	595	9 US-10-072-036-69	Sequence 69, Appl
37	87	7.4	246	9 US-09-933-999A-7	Sequence 7, Appl
38	86	7.4	297	9 US-09-991-211-4	Sequence 4, Appl
39	85	7.3	461	9 US-10-142-835-35	Sequence 35, Appl
40	85	7.3	927	9 US-09-884-696-2	Sequence 2, Appl
41	84	7.2	274	10 US-09-815-242-10868	Sequence 10868, A
42	84	7.2	872	10 US-09-815-242-13594	Sequence 13594, A
43	84	7.2	976	10 US-09-801-368-248	Sequence 248, App
44	82	7.0	317	9 US-10-076-622-176	Sequence 176, App
45	82	7.0	317	9 US-09-551-621-176	Sequence 176, App

ALIGNMENTS

RESULT 1  
US-09-813-453A-12  
; Sequence 12, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OG2-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 12  
; LENGTH: 229  
; TYPE: PRT  
; ORGANISM: Aquifex aeolicus  
US-09-813-453A-12

Query Match				100.0%; Score 1168; DB 9; Length 229;			
Best Local Similarity				100.0%; Pred. No. 1.9e-107;			
Matches 229; Conservative				0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	MRFLTVGVNSSVDIALWEGKKVKDFLKL	SHKEEFKLPKLGISVKSQSFSEKVRGK	60			
Db	1	MRFLTVGVNSSVDIALWEGKKVKDFLKL	SHKEEFKLPKLGISVKSQSFSEKVRGK	60			
Qy	61	IPKIFLKKENFPIQVDYKTPETLGTDRVALAYS	AKKFFGKNVVVISAGTALVIDLVLEG	120			
Db	61	IPKIFLKKENFPIQVDYKTPETLGTDRVALAYS	AKKFFGKNVVVISAGTALVIDLVLEG	120			
Qy	121	KFKGGFITLGLGKKILSDLAEGIPPEPPEEVEIFLGR	STRECVLGGAYRESTEFIKST	180			
Db	121	KFKGGFITLGLGKKILSDLAEGIPPEPPEEVEIFLGR	STRECVLGGAYRESTEFIKST	180			
Qy	181	LKLWRKVKRKKFVWITGGEGKYFSKFGIYDPLLVHGR	MNLLLYLHRI	229			





; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-813-453A-2

Query Match 13.2%; Score 154.5; DB 9; Length 258;
Best Local Similarity 23.7%; Pred. No. 2.5e-07;
Matches 61; Conservative 50; Mismatches 111; Indels 35; Gaps 11;

QY 4 LTVDVGNSSVDIALWEGKKVKDFLKL-SHEEFLKEEFPK-LKAL----GISVKQSFSEKV 57
| :|||: :
Db 3 LVIDVGNTNTVLGVYHDKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGIII 62
:| | | : : | | : : | | : : | | : : | | : : | | : : | |

QY 58 RGKIPKIF----LKKENFPIQ-----VDYKTPETLGTDRVALAYSAKKPYG 100
:| | | : : | | : : | | : : | | : : | | : : | | : : | |
Db 63 SSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNPKYDNPKEVGADRIYNAVAIHLYG 122
:| | | : : | | : : | | : : | | : : | | : : | | : : | |

QY 101 KNVVISAGTALVIDLVLEGG-FKGGFITLGLGKKLKILSDLAEGIPEFFPEEVEIFLGR 159
:| | : ||| : | : | | | : : | | : : | | : : | | : : | |
Db 123 NPLIVVDFGTATTCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIETRPDNIIGK 182
:| : | : | | | | : | | : | | : | | : | | : | | : | |

QY 160 STRECVLGG---AYRESTEFIKSTLKLWRKVKFRKFKVVTGGEGKYFSK-----FGIYDP 212
:| : | : | | | | : | | : | | : | | : | | : | | : | |
Db 183 NTVSAMQSGILFGYGVQVEGIVKRMK-WQA--KODLKVIATGGLAPLIANESDCIDIVDP 239
:| : | : | | | | : | | : | | : | | : | | : | | : | |

QY 213 LLVHRGMRNLLYLXRI 229
| :| : | | :| :
Db 240 FLTLKGL-ELIYERNRV 255
| :| : | | :| :

RESULT 9
US-09-813-453A-49
; Sequence 49, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Bacillus stearothermophilus
US-09-813-453A-49

Query Match 12.7%; Score 148.5; DB 9; Length 258;
Best Local Similarity 23.9%; Pred. No. 9.7e-07;
Matches 60; Conservative 43; Mismatches 111; Indels 37; Gaps 9;

QY 6 VDVGNSSVDIALWEGKKVKDF--LKLSSHEEFLKEEFPKALKALGISVKQSFSEKVRG---- 59
:| | | : : : | : | : | : | : | : | : | : | : | : | : | : |
Db 5 LDVGNTNTVLGVYDGDGLKHHWRIETSRKTEDEYGMIMIKALLNHVGLQFSD-IRGIIIS 63
:| | | : : : | : | : | : | : | : | : | : | : | : | : | : |

QY 60 -KIPKIKFLKKE-----NFPIQVDYKTPETLGTDRVALAYSAKKPYGK 101
:| | | : : | | : : | | : : | | : : | | : : | |
Db 64 SVVPPIMFALERMCLKYFHIKPLIVGPGIKTGLDIKYDNPREVGAADRIYNAVAGIHLYS 123
:| | | : : | | : : | | : : | | : : | | : : | |

QY 102 NVVISAGTALVIDLVLEGG-FKGGFITLGLGKKLKILSDLAEGIPEFFPEEVEIFLGRS 160
:| | : ||| : | : | | | : : | | : : | | : : | | : : | |
Db 124 PLIIVDFGTATTCYINEHKQYMGGAIAPGIMISTEALFARAAKLPRIETARPDIIIGN 183
:| | : ||| : | : | | | : : | | : : | | : : | | : : | |

QY 161 TRECIVLGG---AYRESTEFIKSTLKLWRKVKFRKFKVVTGGEGKYFSK-----FGIYDPL 213
| :| : | | :| | : | | :| | : | | :| | : | | :| |
Db 184 TVSAMQAGILYGVQVEGIVSRMKAKSKI---PPKVIATGGLAPLIASESDIIDVDPF 240
:| : | : | | :| | : | | :| | : | | :| | : | | :| |

QY 214 LVHRGMRNLLY 224
| :| : | | :| |
Db 241 LTLTGLK-LLY 250
| :| : | | :| |

RESULT 10
US-09-813-453A-11
; Sequence 11, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 262
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-09-813-453A-11

Query Match 12.1%; Score 141.5; DB 9; Length 262;
Best Local Similarity 28.3%; Pred. No. 4.9e-06;
Matches 62; Conservative 38; Mismatches 94; Indels 25; Gaps 9;

QY 4 LTVDVGNSSVDIALWEGKKVKDFLK-----LSHEE---FLKE--EFPKALKALGISVKQS 52
| :| | :| : :| | :| : :| | :| : :| | :| : :| | :| :
Db 9 LIIDIGNTSIAFALFKDNQVNLFIKMKTNLMRLRYDEVYSFFEEFNDFNVNKFVFISSVPI 68
:| : | : | | :| : | : | | :| : | : | | :| : | : | |

QY 53 FSEKVRGKIPKIKFLKKE-----NFPIQVD-----YKTPETLGTDRVALAYSAKKPYG-K 101
:| : | : | | :| : | : | | :| : | : | | :| : | : | |
Db 69 LNETFKNVI--FSFFKIKPLEIFGFDLNYDLTFNPKYKSKDKFLGSDVFANLVAAIENYSFE 126
:| : | : | | :| : | : | | :| : | : | | :| : | : | |

QY 102 NVVISAGTALVIDLV-LEGGFKGGFITLGLGKKLKILSDLAEGIPEFFPEEVEIFLGRS 160
||| : ||| : | : | | : | | : | | : | | : | | : | | : | |
Db 127 NVLVVDLGTACTIFAVSRDGIILGGIINSGLPLINFNSLLDNAYLIKFPPISTPNLLERT 186
||| : ||| : | : | | : | | : | | : | | : | | : | | : | |

QY 161 TRECIVLGGAYRESTEFIKSTLKLWRKVKFRKFKVVTGG 199
| | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 187 TSGSVNSGLFYQKYLIEGVYRDIKQMYKKKFNLIITGG 225
| | | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 11
US-09-813-453A-7
; Sequence 7, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:56 ; Search time 6.6898 Seconds  
(without alignments)  
1007.182 Million cell updates/sec

Title: US-09-813-453A-12  
Perfect score: 1168  
Sequence: 1 MRFLTVDVGNSSVDIALWEG.....YDPLLVRGMRNLLLYHRI 229

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

otal number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	86.5	7.4	351	4	US-09-457-040B-6	Sequence 6, Appli
2	86	7.4	297	4	US-09-632-947B-4	Sequence 4, Appli
3	85.5	7.3	350	4	US-09-457-040B-37	Sequence 37, Appli
4	84.5	7.2	872	2	US-08-844-057-2	Sequence 2, Appli
5	84.5	7.2	872	4	US-09-006-730-2	Sequence 2, Appli
6	81.5	7.0	311	4	US-09-175-172-2	Sequence 2, Appli
7	81.5	7.0	1019	4	US-09-434-066-23	Sequence 23, Appli
8	81	6.9	533	4	US-09-603-185-2	Sequence 2, Appli
9	79	6.8	321	2	US-08-922-146-2	Sequence 2, Appli
10	79	6.8	321	2	US-08-922-146-4	Sequence 4, Appli
11	79	6.8	321	4	US-09-343-986-2	Sequence 2, Appli
12	79	6.8	321	4	US-09-343-986-4	Sequence 2, Appli
13	79	6.8	321	4	US-09-322-081-2	Sequence 2, Appli
14	79	6.8	321	4	US-09-322-081-4	Sequence 4, Appli
15	79	6.8	340	1	US-08-097-831-2	Sequence 2, Appli
16	79	6.8	392	4	US-09-026-408-2	Sequence 2, Appli
17	79	6.8	405	4	US-09-026-408-13	Sequence 2, Appli
18	79	6.8	406	1	US-08-434-881-2	Sequence 4, Appli
19	79	6.8	406	3	US-08-977-771-2	Sequence 2, Appli
20	79	6.8	406	4	US-09-361-773-2	Sequence 2, Appli
21	78.5	6.7	247	4	US-08-858-207A-300	Sequence 300, App
22	78	6.7	252	4	US-09-071-035-20	Sequence 20, Appli
23	78	6.7	272	4	US-09-071-035-18	Sequence 18, Appli
24	78	6.7	324	1	US-08-597-236-10	Sequence 10, Appli
25	78	6.7	324	1	US-08-746-682A-10	Sequence 10, Appli
26	77.5	6.6	721	4	US-09-134-078-19	Sequence 19, Appli
27	77	6.6	414	2	US-08-599-171A-25	Sequence 25, Appli

28	77	6.6	414	2	US-08-646-590B-25	Sequence 25, Appli
29	77	6.6	414	3	US-09-069-226-25	Sequence 25, Appli
30	77	6.6	414	4	US-09-412-184-25	Sequence 25, Appli
31	76	6.5	537	3	US-09-028-934-29	Sequence 29, Appli
32	75.5	6.5	459	4	US-09-129-112-9	Sequence 9, Appli
33	75	6.4	511	4	US-08-676-444-42	Sequence 42, Appli
34	75	6.4	857	4	US-08-887-534A-47	Sequence 47, Appli
35	74	6.3	277	1	US-08-400-413-1	Sequence 1, Appli
36	74	6.3	911	4	US-08-461-562B-2	Sequence 2, Appli
37	73.5	6.3	281	4	US-09-134-001C-4575	Sequence 4575, Ap
38	73.5	6.3	383	4	US-09-134-001C-3701	Sequence 3701, Ap
39	73	6.2	776	1	US-08-021-601-2	Sequence 2, Appli
40	73	6.2	776	1	US-08-082-849B-2	Sequence 2, Appli
41	73	6.2	776	5	PCT-US94-01624-2	Sequence 2, Appli
42	73	6.2	1022	1	US-08-271-364A-8	Sequence 8, Appli
43	73	6.2	1022	2	US-08-222-715B-27	Sequence 27, Appli
44	72.5	6.2	509	3	US-08-822-324-8	Sequence 8, Appli
45	72.5	6.2	1009	4	US-09-693-146-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1  
US-09-457-040B-6  
; Sequence 5, Application US/09457040B  
; Patent No. 6387641  
; GENERAL INFORMATION:  
; APPLICANT: Vertex Pharmaceuticals Incorporated  
; APPLICANT: Bellon, Steve  
; TITLE OF INVENTION: Crystallized P38 Complexes  
; FILE REFERENCE: VPI/98-14  
; CURRENT APPLICATION NUMBER: US/09/457,040B  
; CURRENT FILING DATE: 1999-12-08  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 6  
; LENGTH: 351  
; TYPE: PRT  
; ORGANISM: Mouse  
US-09-457-040B-6

Query Match 7.4%; Score 86.5; DB 4; Length 351;  
Best Local Similarity 18.7%; Pred. No. 0.6;  
Matches 53; Conservative 46; Mismatches 74; Indels 111; Gaps 13;

QY	8	VGNSSVDIALWEGKKVKVDFLKLSSHEEFLK-----EPPKLKALG-----	46
Db	1	MGNAAAANKGSEQESVKEFLAKAKEDFLKKWETPSQNTAQLDQDFRIKTLGTGSGFRVML	60
QY	47	ISVKQS---FSEKV--RGKIPKIK-----FLKKENFP-----IQVDYKTPETLGT	87
Db	61	VKHESGNHYAMKILDQKVVKLKQIEHTLNEKRILQAVNFPFLVKLEFSFKDNSNL---	117
QY	83	RVALAYSA-----KKFYG-----KNVVVIS	107
Db	118	YVMVEYVAGGEMFSLRRIGRFSEPHARFYAAQIVLTFFEYLSLDLIYRDLKPENLLIDQ	177
QY	108	AGTALVIDLVLEKFKGGFITL-----GLGKKLK-----ILSDLAEGIP	146
Db	178	QGYIQVTDGFAKRVKGRWTLCGTPEYLAPEIILSKGYNKAVDWALGVLIYEMAAGYP	237
QY	147	EFFPEE-VEIFLGRSTRCEVLGGAYRESTEFIKSTLKLWRKVPK	189
Db	238	PFADQPIQIY-----EKIVSGKVRFPFSHFSSDLKDLLRNLQ	275

RESULT 2  
US-09-632-947B-4  
; Sequence 4, Application US/09632947B  
; Patent No. 6356845  
; GENERAL INFORMATION:  
; APPLICANT: Pharmacia & Upjohn





NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/006,730  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/844,057  
FILING DATE: 18-APR-1997  
APPLICATION NUMBER: 9607991.8  
FILING DATE: 18-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Gimmi, Edward R  
REGISTRATION NUMBER: 38,891  
REFERENCE/DOCKET NUMBER: P31458-5  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 872 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-006-730-2

Query Match 7.2%; Score 84.5; DB 4; Length 872;  
Best Local Similarity 27.7%; Pred. No. 3.3;  
Matches 46; Conservative 25; Mismatches 52; Indels 43; Gaps 10;  
QY 29 LSHEEFLKEEFPKPKALGIVSKQSFSEKVRGKIPKIKFLKKNFPIQVDYKTPETLGTDR 88  
Db 591 LNEVEFLRFDTHFQAVTAELRAIEQQVNEKIWEALEVK---TVETDIDTAKEMG--- 643  
QY 89 VALAYSARKFYGNVVISAGTALVIDLVLE---GKFKGGFITLGLGKKLKILSDLAEGI 145  
Db 644 -AIALFGEK-YGRAVRVTIG-----DYSIELCGGTHVGNLTSEIGLFKIVK-----EEGI 691  
QY 146 PEFFPEEVEIFLGRSTRE--CVLG---GAYRESTEFIK---STLK 182  
Db 692 -----GSGTRRILAVTGKEAFEAYREQEDALKAVAATLK 725  
RESULT 6  
US-09-175-172-2  
; Sequence 2, Application US/09175172  
; Patent No. 6255075  
; GENERAL INFORMATION:  
; APPLICANT: Huang, Jianzhong  
; APPLICANT: McDevitt, Damien  
; TITLE OF INVENTION: bira  
; FILE REFERENCE: GM10172  
; CURRENT APPLICATION NUMBER: US/09/175,172  
; CURRENT FILING DATE: 1998-10-20  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 311  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae

US-09-175-172-2  
Query Match 7.0%; Score 81.5; DB 4; Length 311;  
Best Local Similarity 18.9%; Pred. No. 1.7;  
Matches 54; Conservative 40; Mismatches 67; Indels 125; Gaps 14;  
QY 20 GKKVKDFLKLKLSHEEFLKEEFPKPKALGI---SVKQSFSEKVRGK--IPKIKFLKKNFPI 74  
Db 20 GEKIAEKLSTRTSIWK-AIKRLEQEGEIDSIKNRGYKLMNGDLILPEI---LEENLPI 75  
QY 75 QVDYKTPETLGT-----DRVALAYSA-----KKFYG----- 100  
Db 76 KVSFK-PETKSTQLDAKEAIDLGHANTLYLASYQTAGRGRFORSFYSPQGGIYMTLHLK 134  
QY 101 -----KNVVVISAGTALVIDLVLEGKFKGGFITLGL-GKKLK 136  
Db 135 PNLPYDKLPSYTLVAGAVYKAIKNLTLIDVDIKWVNDIYLNHNHIGGILTEAMTSVETG 194  
QY 137 ILSDLAEGIP-----EFFPEEVEIFLGRSTRECVLGGAYRESTEFIKST-----LK 182  
Db 195 LVTDIIGVGINFITIKDPQELK-----EKAASLFKATAPITRNEIIE 238  
QY 183 LWRKVFRRKFKVVITGGEKGYFSKFGIYDPLLVRHGMNRNLLYLHR 228  
Db 239 IWRAFFE-----TPAE-----ELLYLYKK 257  
RESULT 7  
US-09-434-066-23  
; Sequence 23, Application US/09434066  
; Patent No. 6465714  
; GENERAL INFORMATION:  
; APPLICANT: Luthman, L. Holger  
; APPLICANT: Galli, L.G. Joakim  
; TITLE OF INVENTION: Congenic Animal Models of No. 6465714-Insulin  
; TITLE OF INVENTION: Dependent Diabetes Mellitus  
; FILE REFERENCE: 09705/009001  
; CURRENT APPLICATION NUMBER: US/09/434,066  
; CURRENT FILING DATE: 1999-11-05  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 23  
; LENGTH: 1019  
; TYPE: PRT  
; ORGANISM: Rat  
US-09-434-066-23

Query Match 7.0%; Score 81.5; DB 4; Length 1019;  
Best Local Similarity 23.4%; Pred. No. 8.3;  
Matches 54; Conservative 31; Mismatches 69; Indels 77; Gaps 14;  
QY 32 EEFLKEEFPKPKALGIVSK-----QSFSEKVRGK-----IPKIKFLKKNFPIQVDY 78  
Db 407 QEWVFQECKDLNAVAFRFDKPRGRGYTSKIAGKLHYYPLNGVLTAEYLDLEFRPDLD- 465  
QY 79 KTPETLGTDRVALAYSARKFYGNVVIS-AGT-----ALVIDLV-----LEGKFKGG 125  
Db 465 MVLDKLRPENVRVAIVSKSFEGKTDRTQWYGTQYKQEAIPEDVIQKQWQADLNGKFK-- 523  
QY 126 FITLGLGKKLKILSDLAEGIP---EFFPEEVEIFLGRSTRECVLGGAYRESTEF---IKS 179  
Db 524 -----LPTKNEFIPNTFEIL-----ALEKDATPYPALIKD 553  
QY 180 TL--KLWRKVFRRKFKVVITGG--EGKYFSKFGIYDPLLVRHGMNRNLLYL 226  
Db 554 TMSKLM---FKQDDKFFLPRKACLNFEFFSPFAYVDPLHC-----NMAYLY 596  
RESULT 8  
US-09-603-185-2  
; Sequence 2, Application US/09603185  
; Patent No. 6271004  
; GENERAL INFORMATION:

APPLICANT: Warthoe, Peter  
TITLE OF INVENTION: A METHOD FOR IMPROVED REVERSE TRANSCRIPTION AT HIGH TEMPERATURES  
FILE REFERENCE: 674513-2002  
CURRENT APPLICATION NUMBER: US/09/603,185  
CURRENT FILING DATE: 2000-06-26  
PRIOR APPLICATION NUMBER: DK199900897  
PRIOR FILING DATE: 1999-06-25  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 2  
LENGTH: 533  
TYPE: PRT  
ORGANISM: Pyrococcus sp.  
US-09-603-185-2

Query Match 6.9%; Score 81; DB 4; Length 533;  
Best Local Similarity 21.8%; Pred. No. 3.9;  
Matches 49; Conservative 38; Mismatches 72; Indels 66; Gaps 11;  
QY 4 LTVDVGNSSVDIALWEGKKVKDFLKLKLSHEEFLKEEFPKPKALG---ISVKQSFSEKVRGK 60  
Db 246 LEVKETETDAEIRITSPEQLQAFLE-QEEKMLREVMVDKKEVGANVVFQKGIDDLAQHY 304  
QY 61 IPK-----IKFLKKENFP-----IQVDYKTPETLGTDRVALAYSAKKFGYKNV- 103  
Db 305 LAKYGIMAVRRVKKSDMEKLAKATGAKIVTNVRDLTPEDLGE--AELVEQKRVAGENMI 361  
QY 104 -----VVISAGTALVIDLVLEGKFGGFFITLGLGKKLKLSDLAE----- 143  
Db 362 FVEGCKNPKAVTILIRGGTEHVDEV-----ERALEDVAVKVKDIVEDGKIVAAG 411  
QY 144 GIPEFFPEEVEIFLGRSTRECVLGG-----AYRESTEEFKSTL 181  
Db 412 GAPEI---ELAIRLDEYAKE--VGGKEQLAIEAFAEALKVIPRTL 451

RESULT 9  
US-08-922-146-2  
Sequence 2, Application US/08922146  
Patent No. 5965402  
GENERAL INFORMATION:  
APPLICANT: Burnham, Martin K. R.  
APPLICANT: Lonetto, Michael A.  
APPLICANT: Warren, Patrick V.  
TITLE OF INVENTION: No. 5965402el phoH Homolog  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dechert, Price & Rhoads  
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103-2793  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSEQ for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/922,146  
FILING DATE: 25-AUG-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION NUMBER:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson, Arthur E  
REGISTRATION NUMBER: 34,354  
REFERENCE/DOCKET NUMBER: P50444-03  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 609/520-3254  
TELEFAX: 609/520-3259  
TELEX:

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 321 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-922-146-2  
Query Match 6.8%; Score 79; DB 2; Length 321;  
Best Local Similarity 20.7%; Pred. No. 3.1;  
Matches 51; Conservative 42; Mismatches 105; Indels 48; Gaps 10;  
QY 6 VDVGNS-----SVDIALWEGKKVKDFLKLKLSHEEFLKEEFPKPKALGISVKQSFSEKVRG 59  
Db 72 IDLGNNTIKDVEAAIKMAHNNTIQHLLLDLYDEEITKDAFGK-----TIRAKTMG 121  
QY 60 KIPKIKFLKKENFPQVDYKTPETLGTDRVALAYSAK---KFGYKNVV-----ISAGTAL 112  
Db 122 QRIYVNAKNNNDLVFGIG---PAGTGKTFLLAVVYAAKQLRKGAGKRIVLTPRAVEAGESL 178  
QY 113 -----VIDLVLEGKFGGFFITLGLGKKLKLSDLAEGIPPEEVEIFLGRSTRE 163  
Db 179 GFLPGDLKEKVDPLYRLPYLDGLYTVLGREQTERFIE---RGIIEIAP--LAYMRGRTLED 233  
QY 164 CVLGGAYRESTEFIKSTLKLWRKVKFKPKVVTGGEGKYFSKFGIYDPLLVHRGMNLL 223  
Db 234 AFVILDEAQNTHAQMFMFLTRLGFGS--KMVVTGDQTO-----IDLPGKVGKSGLKEAV 285  
QY 224 YLYHRI 229  
Db 286 SRLHNV 291

RESULT 10  
US-08-922-146-4  
Sequence 4, Application US/08922146  
Patent No. 5965402  
GENERAL INFORMATION:  
APPLICANT: Burnham, Martin K. R.  
APPLICANT: Lonetto, Michael A.  
APPLICANT: Warren, Patrick V.  
TITLE OF INVENTION: No. 5965402el phoH Homolog  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dechert, Price & Rhoads  
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103-2793  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSEQ for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/922,146  
FILING DATE: 25-AUG-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION NUMBER:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson, Arthur E  
REGISTRATION NUMBER: 34,354  
REFERENCE/DOCKET NUMBER: P50444-03  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 609/520-3254  
TELEFAX: 609/520-3259  
TELEX:  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 321 amino acids



;;  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/322,081  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA: US/08/922,146  
;; APPLICATION NUMBER: US/08/922,146  
;; FILING DATE: 25-AUG-1997  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Jackson, Arthur E  
;; REGISTRATION NUMBER: 34,354  
;; REFERENCE/DOCKET NUMBER: P50444-03  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 609/520-3254  
;; TELEFAX: 609/520-3259  
;; TELEX:  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 321 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; US-09-322-081-2

Query Match 6.8%; Score 79; DB 4; Length 321;  
Best Local Similarity 20.7%; Pred. No. 3.1;  
Matches 51; Conservative 42; Mismatches 105; Indels 48; Gaps 10;  
QY 6 VDVGNS-----SVDIALWEGKKVKDFLKLSSHEEFLKEEFPKALKGKISVKQSFSEKVRG 59  
Db 72 IDLGNNTIKDVEAAIKMAHNNTIQHLLDLYDEEITKDAFGK-----TIRAKTMG 121  
QY 60 KIPKIKFLKKENFPIQVDYKTPETLGTDRVALAYSAK---KFYGNVVV-----ISAGTAL 112  
Db 122 QRIYVNAMKNDLVFGIG---PAGTGKTF LAVVYAAQLRKAGKGRIVLTRPAVEAGESL 178  
QY 113 -----VIDLVLEGKFGGFTLGLGKKLKILSDLAEGIPPEFFPEEVEIFLGRSTRE 163  
Db 179 GFLPGDLKEKVDPYLRPLYDGLYTVLGRQTERFIE----RGIIEIAP--LAYMRGRTLED 233  
QY 164 CVLGGAYRESTEFIKSTLKLWRKVKRKFVVITGGEGKYFSKFGIYDPLLVHVRGMRNLL 223  
Db 234 AFVILDEAQNTHAQMFMFLTRLGFGS--KMVVTGDQTO-----IDLPGVKSGLKEAV 285  
QY 224 YLYHRI 229  
Db 286 SRLHNV 291

RESULT 14  
;-09-322-081-4  
; Sequence 4, Application US/09322081  
; Patent No. 632336  
; GENERAL INFORMATION:  
; APPLICANT: Burnham, Martin K. R.  
; APPLICANT: Lonetto, Michael A.  
; APPLICANT: Warren, Patrick V.  
; TITLE OF INVENTION: No. 6323336el phoH Homolog  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dechert, Price & Rhoads  
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103-2793  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows  
; SOFTWARE: FastSeq for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/322,081  
; FILING DATE:

;;  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/922,146  
;; FILING DATE: 25-AUG-1997  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Jackson, Arthur E  
;; REGISTRATION NUMBER: 34,354  
;; REFERENCE/DOCKET NUMBER: P50444-03  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 609/520-3254  
;; TELEFAX: 609/520-3259  
;; TELEX:  
;; INFORMATION FOR SEQ ID NO: 4:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 321 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; US-09-322-081-4

Query Match 6.8%; Score 79; DB 4; Length 321;  
Best Local Similarity 20.7%; Pred. No. 3.1;  
Matches 51; Conservative 42; Mismatches 105; Indels 48; Gaps 10;  
QY 6 VDVGNS-----SVDIALWEGKKVKDFLKLSSHEEFLKEEFPKALKGKISVKQSFSEKVRG 59  
Db 72 IDLGNNTIKDVEAAIKMAHNNTIQHLLDLYDEEITKDAFGK-----TIRAKTMG 121  
QY 60 KIPKIKFLKKENFPIQVDYKTPETLGTDRVALAYSAK---KFYGNVVV-----ISAGTAL 112  
Db 122 QRIYVNAMKNDLVFGIG---PAGTGKTF LAVVYAAQLRKAGKGRIVLTRPAVEAGESL 178  
QY 113 -----VIDLVLEGKFGGFTLGLGKKLKILSDLAEGIPPEFFPEEVEIFLGRSTRE 163  
Db 179 GFLPGDLKEKVDPYLRPLYDGLYTVLGRQTERFIE----RGIIEIAP--LAYMRGRTLED 233  
QY 164 CVLGGAYRESTEFIKSTLKLWRKVKRKFVVITGGEGKYFSKFGIYDPLLVHVRGMRNLL 223  
Db 234 AFVILDEAQNTHAQMFMFLTRLGFGS--KMVVTGDQTO-----IDLPGVKSGLKEAV 285  
QY 224 YLYHRI 229  
Db 286 SRLHNV 291

RESULT 15  
US-08-097-831-2  
; Sequence 2, Application US/08097831  
; Patent No. 5510473  
; GENERAL INFORMATION:  
; APPLICANT: Camerini-Otero, Rafael D.  
; APPLICANT: Angov, Evangelina  
; TITLE OF INVENTION: Cloning and Expression of Taq reca  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: KNOBBE, MARTENS, OLSON AND BEAR  
; STREET: 620 Newport Center Drive 16th Floor  
; CITY: Newport Beach  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/097,831  
; FILING DATE: 19930726  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Altman, Daniel E.  
; REGISTRATION NUMBER: 34,115



REFERENCE/DOCKET NUMBER: NIH066.001A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 714-760-0404  
TELEFAX: 714-760-9502  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 340 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-097-831-2

Query Match 6.8%; Score 79; DB 1; Length 340;  
Best Local Similarity 24.2%; Pred. No. 3.4;  
Matches 40; Conservative 25; Mismatches 50; Indels 50; Gaps 7;  
QY 66 FLKKENFPIQVDYKTPETLGTDRVALAYSAKKFGKNNVVVISAGTALVIDLVLEK-FKG 124  
-Db 189 FINQVREKVGVMYGNPETTPGGR-----ALKFY-----SSVRLDVRKSGQPIKV 232  
/ 125 GFITLGLGKKLKILSD-LAEGIPPEFFPEEVEIFLGR----- 159  
Db 233 GNEAVGKIKVKVKNKLA---PPFREAELEIYFGRGLDPVMDLVNVVAAGVIEKAGSW 289  
QY 160 -STRECVLGAYRESTEFIKSTLKLWR-----KVFKRKFKVITGG 199  
Db 290 FSYGEHRLGQCKEKAAYLRRERPELLEIRAKVLERADKVVLAAG 334

Search completed: June 24, 2003, 21:49:26  
Job time : 7.6898 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.  
OM protein - protein search, using sw model  
Run on: June 24, 2003, 21:35:06 ; Search time 18.6773 seconds  
(without alignments)  
2526.317 Million cell updates/sec

Title: US-09-813-453A-12  
Perfect score: 1168  
Sequence: 1 MRFLTVDVGNSSVDIALWEG.....YDPLLVHRGMNLLLYHRI 229  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 671580 seqs, 206047115 residues  
total number of hits satisfying chosen parameters: 671580  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPREMBL\_21:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvirus:\*

16: sp\_bacteriap:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1168	100.0	229	16	O67753 aquifex ao
2	193.5	16.6	246	16	Q9WZY5 thermotoga
3	177.5	15.2	209	16	Q9PIA9 campylobact
4	156.5	13.4	259	16	Q8YAC5 listeria mo
5	149.5	12.8	259	16	Q92F54 listeria in
6	145	12.4	256	16	Q8RFE4 fusobacteri
7	141.5	12.1	262	16	O51477 borrellia bu
8	138.5	11.9	223	16	O25533 helicobacte
9	138.5	11.9	254	16	Q9KGH5 bacillus ha
10	138.5	11.9	273	16	Q97EB4 clostridium
11	137.5	11.8	258	2	Q9F985 bacillus st
12	136.5	11.7	255	16	Q8R7M2 thermoanaer
13	133.5	11.4	262	16	Q9RX54 deinococcus
14	133	11.4	257	16	P74045 synechocyst
15	131.5	11.3	223	16	Q9ZKY6 helicobacte
16	129.5	11.1	261	16	Q9A6Z1 caulobacter

17	128.5	11.0	265	16	Q9X8N6	Q9x8n6 streptomyc
18	122.5	10.5	273	16	O83446	O83446 treponema p
19	120.5	10.3	212	2	O32514	O32514 desulfovibr
20	120.5	10.3	259	16	O8XHL5	O8xhl5 clostridium
21	117	10.0	276	16	O8YQD7	O8yqd7 anabaena sp
22	114	9.8	272	16	O06282	O06282 mycobacteri
23	113.5	9.7	295	16	O8Y2M4	O8y2m4 ralstonia s
24	112	9.6	274	16	O9CD56	O9cd56 mycobacteri
25	108	9.2	592	16	O9JWI7	O9jwi7 neisseria m
26	105	9.0	224	16	O98Q93	O98q93 mycoplasma
27	103	8.8	592	16	O9JXF1	O9jxf1 neisseria m
28	102.5	8.8	254	17	O28311	O28311 archaeoglob
29	101.5	8.7	476	16	O8RH37	O8rh37 fusobacteri
30	100.5	8.6	763	16	O8XJY4	O8xjy4 clostridium
31	98.5	8.4	248	16	O9HWC1	O9hwc1 pseudomonas
32	97.5	8.3	484	16	O9ZMF7	O9zmf7 helicobacte
33	96.5	8.3	328	10	O9XQB0	O9xqb0 phaseolus a
34	96	8.2	543	5	O9U4D6	O9u4d6 oxytricha n
35	96	8.2	578	16	O97H45	O97h45 clostridium
36	95.5	8.2	426	17	O57987	O57987 pyrococcus
37	95	8.1	242	16	O9PCI4	O9pci4 xylella fas
38	95	8.1	390	5	O8T8V5	O8t8v5 drosophila
39	95	8.1	816	16	O8RB00	O8rb00 thermoanaer
40	94.5	8.1	484	16	O25055	O25055 helicobacte
41	94	8.0	483	8	O9GER7	O9ger7 omphalogram
42	93.5	8.0	427	3	O9URY5	O9ury5 schizosacch
43	93.5	8.0	613	16	O8RDP8	O8rdp8 fusobacteri
44	93	8.0	267	12	O8V9A4	O8v9a4 crucifer to
45	93	8.0	487	8	O9GEQ9	O9geq9 soldanella

ALIGNMENTS

RESULT 1			
O67753	PRELIMINARY;	PRT;	229 AA.
AC	O67753;		
DT	01-AUG-1998 (TrEMBLrel. 07, Created)		
DT	01-AUG-1998 (TrEMBLrel. 07, Last sequence update)		
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)		
DE	Hypothetical protein AQ_1924.		
GN	AQ_1924.		
OS	Aquifex aeolicus.		
OC	Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.		
OX	NCBI_TaxID=63363;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=VF5;		
RX	MEDLINE=98196666; PubMed=9537320;		
RA	Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,		
RA	Graham D.E., Overbeek R., Snead M.A., Keller M., Aufay M., Huber R.,		
RA	Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;		
RT	"The complete genome of the hyperthermophilic bacterium Aquifex		
RT	aeolicus.";		
RL	Nature 392:353-358(1998).		
DR	EMBL; AE000763; AAC07720.1; -.		
DR	InterPro; IPR004619; Baf.		
DR	Pfam; PF03309; Bvg_acc_factor; 1.		
DR	TIGRFAMS; TIGR00671; baf; 1.		
KW	Hypothetical protein; Complete proteome.		
SQ	SEQUENCE 229 AA; 26068 MW; 64C347A285FABE9A CRC64;		
Query Match 100.0%; Score 1168; DB 16; Length 229;			
Best Local Similarity 100.0%; Pred. No. 5.1e-87;			
Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1	MRFLTVDVGNSSVDIALWEGKKVKDFLKLKSHEEFLKEEFPKIKALGISVKQSFSEKVRGK	60
Db	1	MRFLTVDVGNSSVDIALWEGKKVKDFLKLKSHEEFLKEEFPKIKALGISVKQSFSEKVRGK	60
QY	61	IPKIKFLKKENFPIQVDYKTPETLTGTRVALAYSAKKFYGNVVVISAGTALVIDLVLEG	120

Db 61 IPKIKFLKKENFPIQVDYKTPETLGTDRVALAYSAKKFGKNNVVWISAGTALVIDLVLEG 120

QY 121 KFKGGFITLGLGKKILSDLAEGIPPEFFPEEVEIFLGRSTRECVLGGAYRESTEFIKST 180

Db 121 KFKGGFITLGLGKKILSDLAEGIPPEFFPEEVEIFLGRSTRECVLGGAYRESTEFIKST 180

QY 181 LKLWRKVKFRKFKVVTGGGKGYFSKFGIYDPLLVHVRGNLLYLYHRI 229

Db 181 LKLWRKVKFRKFKVVTGGGKGYFSKFGIYDPLLVHVRGNLLYLYHRI 229

RESULT 2

Q9WZY5

ID Q9WZY5 .PRELIMINARY; PRT; 246 AA.

AC Q9WZY5;

DT 01-NOV-1999 (TREMBlrel. 12, Created)

DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Hypothetical protein TM0883.

GN TM0883.

OS Thermotoga maritima.

OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.

OX NCBI\_TaxID=2336;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MSB8 / DSM 3109;

RX MEDLINE=99287316; PubMed=10360571;

RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A., McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M., Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D., Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O., Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;

RA "Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima.";

RT Nature 399:323-329(1999).

RL EMBL; AE001754; AAD35964.1; -.

DR TIGR; TM0883;

DR InterPro; IPR004619; Baf.

DR Pfam; PF03309; Bvg\_acc\_factor; 1.

DR TIGRFAMs; TIGR00671; baf; 1.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 246 AA; 27154 MW; 9E0309AD462CF266 CRC64;

Query Match 16.6%; Score 193.5; DB 16; Length 246;

Best Local Similarity 26.5%; Pred. No. 6e-08;

Matches 66; Conservative 45; Mismatches 105; Indels 33; Gaps 6;

Y 3 FLTVDVGNSSVDIALWEGKKVKDFLKLKSHEEFLKKE-----FPKLKALGI-- 47

Y 2 YLLVDVGNTHSVFSITEDGKTFRRWRRLSTGVFQTEDELFSHLHPLLDGAMREIKGIGVAS 61

QY 48 -----SVKQSFSEKVRGKIPKIKFLKKENFPIQVDYKTPETLGTDRVALAYSAKKFGYK 101

Db 62 VVPTQNTVIERFSQKYFHISP--IWVKAKNGCVKNNVKNPSEVGADRVANVVFVKEYGK 119

QY 102 NVVVISAGTALVIDLVLECKFKGGFITLGLGKKILSDLAEGIPPEFFPEEVEIFLGRST 161

Db 120 NGIIDMGATTVDLVVNGSYEGGAILPGFFMMVHSLFRGTAKLPLVEVKPADVVGKDT 179

QY 162 RECVLGGAYRESTEFIKSTLKLWRKVKFRKFKVVTGGGKGYFS---KFGIYDPLLVHVRG 218

Db 180 EENIRLGVVNGSVVALEGIIGRIKEY-GDLPVVLTTGGQSKIVKDMIKHEIFDEDLTIKG 238

QY 219 MRNLLYLYH 227

Db 239 -----VYH 241

RESULT 3

Q9PIA9

ID Q9PIA9 .PRELIMINARY; PRT; 209 AA.

AC Q9PIA9;

DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Hypothetical protein Cj0394c.

GN Cj0394c.

OS Campylobacter jejuni.

OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;

OC Campylobacter.

OX NCBI\_TaxID=197;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NCTC 11168;

RX MEDLINE=20150912; PubMed=10688204;

RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C., Basham D., Chillingworth T., Davies R.M., Feltwell T., Holroyd S., Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W., Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M., Whitehead S., Barrell B.G.;

RA "The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervariable sequences.";

RT Nature 403:665-668(2000).

RL EMBL; AL139075; CAB74230.1; -.

DR InterPro; IPR004619; Baf.

DR Pfam; PF03309; Bvg\_acc\_factor; 1.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 209 AA; 23935 MW; C49313AA9E3D2D19 CRC64;

Query Match 15.2%; Score 177.5; DB 16; Length 209;

Best Local Similarity 25.8%; Pred. No. 9.7e-07;

Matches 62; Conservative 44; Mismatches 83; Indels 51; Gaps 10;

QY 4 LTVDVGNSSVDIALWEGK----KVKDFLKLKSHEEFLKKEFPKLKALGISVKQSFSEKVRG 59

Db 2 LLCDIGNSNANF-LDDNKYFTLNIDQFLEFKNEQ-----KIFYINVNEHLKHEHLK- 50

QY 60 KIPKIKFLKKENFPIQVDYKTPET-----LGTDRVALAYSAKKFGKNNVVWISAGTALVID 115

Db 51 -----NQKNFINLEPYFLEFDITYQGLGIDRIAACYTI-----EDGVVVDAGSAITID 97

QY 116 LVLECKFKGGFITLGLGKKILSDLAEGIPPEFFPEEVEI-FLGRSTRECVLGGAYREST 174

Db 98 IISNSIHLGGFILPGIANKYKIIYSHISPRLKSEFNTQVSLDAFPQKTMDSYGVFKGIY 157

QY 175 EPIKSTLKLWRKVKFRKFKVVTGGGKGYFSKF---GIYDPLLVHVRGMR-----NLLY 224

Db 158 LLIKDA-----AQNKLYFTGGDGGQFLANYFDHAIYDKLLIFRGMKKIIKENPNLLY 209

RESULT 4

Q8YAC5

ID Q8YAC5 .PRELIMINARY; PRT; 259 AA.

AC Q8YAC5;

DT 01-MAR-2002 (TREMBlrel. 20, Created)

DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Hypothetical protein lmo0221.

GN LMO0221.

OS Listeria monocytogenes.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;

OC Listeriaceae; Listeria.

OX NCBI\_TaxID=1639;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=EGD-E / SEROVAR 1/2A;

RX MEDLINE=21537279; PubMed=11679669;

RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A., Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T., Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P., Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O., Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P., Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D., Jones L.-M., Kaerst U., Krefit J., Kuhn M., Kunst F., Kurapkat G., Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,















**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model  
Run on: June 24, 2003, 21:33:11 ; Search time 3.98294 Seconds  
(without alignments)  
2384.688 Million cell updates/sec

Title: US-09-813-453A-12  
Perfect score: 1168  
Sequence: 1 MRFLTVDVGNSSVDIALWEG.....YDPLLVRGMRNLLYHRI 229

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues  
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	117	10.0	233	1	YACB_BACSU	P37564 bacillus su
2	101	8.6	596	1	UUP_BUCAI	P57445 buchnera ap
3	100.5	8.6	733	1	K6AA_XENLA	P10665 xenopus lae
4	98	8.4	491	1	RNG_HAEIN	P45175 haemophilus
5	96	8.2	390	1	LA_DROME	P40796 drosophila
6	92	7.9	194	1	TRPG_HELPJ	Q9zju6 helicobacte
7	91	7.8	265	1	MOVP_CRMV	Q66221 chinese rap
8	91	7.8	463	1	NAOX_METJA	Q58065 methanococc
9	90.5	7.7	379	1	YO11_MOUSE	P11260 mus musculu
10	90	7.7	370	1	DCUP_SCHPO	Q9usj5 schizosacch
11	89	7.6	799	1	HIS2_YEAST	P00815 saccharomyc
12	88.5	7.6	480	1	SYFA_METJA	Q57911 methanococc
13	88	7.5	213	1	RS6E_SULTO	Q975n7 sulfolobus
14	87	7.4	397	1	KAPI_BOVIN	P24256 bos taurus
15	87	7.4	593	1	CC23_SCHPO	O42709 schizosacch
16	86.5	7.4	194	1	TRPG_HELPY	O25868 helicobacte
17	86.5	7.4	629	1	K6AB_XENLA	P10666 xenopus lae
18	86.5	7.4	630	1	YO19_METJA	Q60325 methanococc
19	86.5	7.4	708	1	EFLG_SPRCO	P40173 streptomyc
20	86.5	7.4	1176	1	SLAP_BACSH	P38537 bacillus sp
21	86	7.4	265	1	PANB_AQUAE	O67783 aquifex ae
22	86	7.4	297	1	MURB_AQUAE	O66805 aquifex ae
23	86	7.4	461	1	HMCS_ARATH	P54873 arabidopsis
24	85.5	7.3	326	1	SPI_BACBR	P43131 bacillus br
25	85.5	7.3	350	1	KAPA_MOUSE	P05132 mus musculu
26	85	7.3	483	1	MURC_BUCAP	O51926 buchnera ap
27	84	7.2	267	1	MOVP_TVCV	Q88921 turnip vein
28	84	7.2	340	1	RECA_THEAQ	P48296 thermus aqu
29	84	7.2	621	1	GIDA_RICPR	Q9ze90 rickettsia
30	84	7.2	976	1	PDR3_YEAST	P33200 saccharomyc
31	83.5	7.1	248	1	YA14_SCHPO	Q09686 schizosacch
32	83.5	7.1	441	1	DCTA_RHIME	P20672 rhizobium m
33	83	7.1	144	1	HBAL_ONCMY	P02019 oncorhynchu

34	82.5	7.1	848	1	AMPN_LACLA	Q48656 lactococcus
35	82	7.0	251	1	Y7SR_METFE	P21336 methanother
36	82	7.0	448	1	Y861_METJA	Q58271 methanococc
37	82	7.0	546	1	THSB_THEK1	O24730 thermococcu
38	81.5	7.0	255	1	LACR_LACLA	P18816 lactococcus
39	81.5	7.0	350	1	KAPB_BOVIN	P05131 bos taurus
40	81.5	7.0	350	1	KAPB_HUMAN	P22694 homo sapien
41	81.5	7.0	358	1	ID12_LISMO	Q8y7a5 listeria mo
42	81.5	7.0	1019	1	IDE_MOUSE	Q9jhr7 mus musculu
43	81.5	7.0	1019	1	IDE_RAT	P35559 rattus norv
44	81	6.9	365	1	Y075_METJA	Q60388 methanococc
45	81	6.9	546	1	THSB_PYRKO	Q52500 pyrococcus

ALIGNMENTS

RESULT 1  
YACB\_BACSU  
ID YACB\_FACSU STANDARD; PRT; 233 AA.  
AC P37564;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein yacB.  
GN YACB.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=96051385; PubMed=7584024;  
RA Ogasawara N., Nakai S., Yoshikawa H.;  
RT "Systematic sequencing of the 180 kilobase region of the Bacillus  
subtilis chromosome containing the replication origin.";  
RL DNA Res. 1:1-14(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,  
Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
Joris B., Karamata D., Kasahara Y., Klaubner-Blanchard M., Klein C.,  
Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
Lee S.M., Levine A., Liu H., Masuda S., Mael C., Medigue C.,  
Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,  
Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
Viari A., Wambutt R., Wedler E., Wedler H., Weltzenegger T.,  
Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
subtilis.";  
RL Nature 390:249-256(1997).

-1- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.





DR Pfam; PF00433; pkinase\_C; 1.  
DR ProDom; PD000001; Euk\_pkinase; 2.  
DR SMART; SM00133; S\_TK\_X; 1.  
DR SMART; SM00220; S\_TKC; 2.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 2.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 2.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 2.  
KW transferase; Serine/threonine-protein kinase; ATP-binding;  
KW Repeat; Multigene family.  
FT DOMAIN 62 321 PROTEIN KINASE 1.  
FT DOMAIN 416 673 PROTEIN KINASE 2.  
FT NP\_BIND 68 76 ATP (BY SIMILARITY).  
FT BINDING 94 94 ATP (BY SIMILARITY).  
FT ACT\_SITE 187 187 BY SIMILARITY.  
FT NP\_BIND 422 430 ATP (BY SIMILARITY).  
FT BINDING 445 445 ATP (BY SIMILARITY).  
FT ACT\_SITE 533 533 BY SIMILARITY.  
SQ SEQUENCE 733 AA; 82639 MW; 2BA4B137EE7DBDDC CRC64;  
  
Query Match 8.6%; Score 100.5; DB 1; Length 733;  
Best Local Similarity 26.7%; Pred. No. 1.7;  
Matches 63; Conservative 30; Mismatches 74; Indels 69; Gaps 15;  
  
QY 19 EKG--KVKDFL-----KLSHEEFLKEEFPK-----LKALGISVKQSFSEKV 57  
Db 134 EGKLYLIDFLRGDLFTRLSKFEVMTTEEDVKFYLAELALGLDHLHSLGIYRLKPENI 193  
  
QY 58 ----RGKIPKIF-LKKE-----NFPIQVDYKTPETLGTDRVALAYS AKKF-YGKN 102  
Db 194 LLDEEGHKILTDGLSKAEIDHEKKAYSFCGTVEYMAPEVV--NRQGHSHSADWWSYGVL 251  
  
QY 103 VVVISAGTALVIDLVLEGGKFGGFTLGLGKKLILSDLAEGIPPEEFPPEEVE-----IFL 157  
Db 252 MFEMLTGS-----LPFQGKDRKETMTLILKAKL-----GMPQFLSNEAQSLLRALFK 298  
  
QY 158 GRSTRECVLGAYRESTEFIK-----STLKLWRKVFKRK-----FKVVITGGEGKYF 204  
Db 299 RNPTNRL---GSAMEGAEIKRQPFSTID-WNKLFRRMSPFPKPAVTQADDTYY 350

RESULT 4  
RNG\_HAEIN STANDARD; PRT; 491 AA.  
ID RNG\_HAEIN STANDARD; PRT; 390 AA.  
AC P45175;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Ribonuclease G (EC 3.1.4.-) (RNase G) (Cytoplasmic axial filament protein).  
GN CAFA OR RNG OR HI1353.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Rd / KW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M., McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D., Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M., Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D., Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C., Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M., Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.";  
RL Science 269:496-512(1995).  
CC -1- FUNCTION: INVOLVED IN PROCESSING OF THE 5'END OF 16S RNA. COULD BE INVOLVED IN CHROMOSOME SEGREGATION AND CELL DIVISION. IT MAY BE ONE OF THE COMPONENTS OF THE CYTOPLASMIC AXIAL FILAMENTS BUNDLES

CC OR MERELY REGULATE THE FORMATION OF THIS STRUCTURE (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: BELONGS TO THE RNE FAMILY.  
CC -1- SIMILARITY: CONTAINS 1 S1 MOTIF DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U32814; AAC23000.1; -.  
DR TIGR; HI1353; -.  
DR InterPro; IPR004659; RNaseEG.  
DR InterPro; IPR003029; S1.  
DR Pfam; PF00575; S1; 1.  
DR SMART; SM00316; S1; 1.  
DR TIGRFAMS; TIGR00757; RNaseEG; 1.  
DR PROSITE; PS50126; S1; 1.  
KW Hydrolase; Nuclease; Endonuclease; Cell division; Cytoskeleton;  
KW Complete proteome.  
FT DOMAIN 40 129 S1 MOTIF.  
SQ SEQUENCE 491 AA; 55928 MW; 3CFD0E9ACB39FA37 CRC64;  
  
Query Match 8.4%; Score 98; DB 1; Length 491;  
Best Local Similarity 24.0%; Pred. No. 1.6;  
Matches 52; Conservative 33; Mismatches 76; Indels 56; Gaps 8;  
  
QY 13 VDIALWEGKKVKDFLKLKLSHEEFLKEEFPKALGISVKQSFSEKVRGKIPKIKFLKKENF 72  
Db 59 VDIGLEKAAFLHAADIVSHTECVDE-----NEQKQFKVKSISELVREGQ 102  
  
QY 73 PIQVDYKTPETLGTDRVALAYS AKKFYKNNVVVISAGTALVIDLVLEGG---FKGGFITL 129  
Db 103 DIVVQV-VKEPLGT-----KGARLTDTITLPSRHLVFMPENSHV 140  
  
QY 130 GLGKKLILSDLA--EGIPFPEEVEIFLGRSTRECVLGAYRESTEFIKSTLKLWRKV 187  
Db 141 GVSQRIESEERARLKVLPFCDELGGFIIRTEGASEEELRQDAEFLK---RLWRKV 197  
  
QY 188 FKRKFVVITGGEGKYFSKFGIY-DPLLVRHGMNLL 223  
Db 198 LERK-----SKYPTKSKIYGEALPQRIILRDFI 225  
  
RESULT 5  
LA\_DROME STANDARD; PRT; 390 AA.  
ID LA\_DROME STANDARD; PRT; 390 AA.  
AC P40796; Q24375; Q9VIN2;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE La protein homolog (La ribonucleoprotein) (La autoantigen homolog).  
GN LA OR CG10922.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Canton-S; TISSUE=Ovary;  
RX MEDLINE=94309632; PubMed=8035794;  
RA Bai C., Li Z., Tolias P.P.;  
RT "Developmental characterization of a Drosophila RNA-binding protein homologous to the human systemic lupus erythematosus-associated La/SS-B autoantigen.";  
RL Mol. Cell. Biol. 14:5123-5129(1994).  
RN [2]  
RP SEQUENCE FROM N.A.



CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; AE001546; AAD06781.1; -  
DR HSSP; Q06129; IQDL.  
DR InterPro; IPR002385; Anth\_synthII.  
DR InterPro; IPR001317; CPS\_GATase.  
DR InterPro; IPR000991; GATase\_1.  
DR Pfam; PF00117; GATase; 1.  
DR PRINTS; PR00097; ANTSNTHASEII.  
DR PRINTS; PR00099; CPSCATASE.  
DR PRINTS; PR00096; GATASE.  
DR TIGRFAMS; TIGR00566; trpG\_papa; 1.  
DR PROSITE; PS00442; GATASE\_TYPE\_1; 1.  
KW Tryptophan biosynthesis; Lyase; Glutamine amidotransferase;  
KW Complete proteome.  
--FT ACT\_SITE 84 BY SIMILARITY.  
T ACT\_SITE 170 BY SIMILARITY.  
T ACT\_SITE 172 BY SIMILARITY.  
SQ SEQUENCE 194 AA; 21479 MW; 47A123C678DBE119 CRC64;

Query Match 7.9%; Score 92; DB 1; Length 194;  
Best Local Similarity 26.8%; Pred. No. 1.7;  
Matches 42; Conservative 18; Mismatches 63; Indels 34; Gaps 6;

QY 35 LKEEFPKLGKALGIVSVKQSFSEKVRGKIPKIKLKKENFPIQVDYKTPETLGTDRVALAYS 94  
| | | | | : | | | | | : | | | | | : | | : | |  
Db 42 LMNEEPKTPLLFISPGPG-NPNSSGNLLKIIMAKKKFPI-----LGVCILGLQALQSYG 95  
  
QY 95 AKFYGKNVVVISAGTALVIDLVLEGFKG-----GFITLGLGKKLKILSDLA 143  
| | | | | : | | | | | : | | | | | : | | : | |  
Db 96 AKIIRSKEIV---HGKATTIALKKHAFVFKGLGSEMVVGRYHSLMASGLPKNLEVIAE-HD 151  
  
QY 144 GIP-----EFFPEEVEIFLGRSTRECVLG 167  
| | : | | : | | : | | : | | : | | : | | : | |  
Db 152 NIPMAIVNEEDKILAYQFHPESIMTLQGRALLEQSVG 188

RESULT 7  
MOV\_P\_CRMV STANDARD; PRT; 265 AA.  
AC Q66221;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Movement protein (Cell-to-cell transport protein) (30 kDa protein).  
GN MP.  
OS Chinese rape mosaic virus (CRMV) (Oilseed rape mosaic virus).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.  
OX NCBI\_TaxID=42007;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96197410; PubMed=8616237;  
RA Aguilar I., Sanchez F., Martin-Martin A., Martinez-Herrera D.,  
RA Ponz F.;  
RT "Nucleotide sequence of Chinese rape mosaic virus (oilseed rape mosaic  
RT virus), a crucifer tobamovirus infectious on Arabidopsis thaliana.";  
RL Plant Mol. Biol. 30:191-197(1996).  
CC -1- FUNCTION: Involved in transport of the virus from the initially  
CC infected cells to adjacent cells, possibly by modifying the  
CC function of the plasmodesmata. Also influences local lesion  
CC development. Binds to RNA and single-stranded DNA.  
CC -1- SIMILARITY: BELONGS TO THE TOBAMOVIRUSES MOVEMENT PROTEIN FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

CC -----  
DR EMBL; U30944; AAB60600.1; -  
DR InterPro; IPR001022; TMV\_movement.  
DR Pfam; PF01107; Tobamo\_MP; 1.  
DR PRINTS; PR00964; MOVEMENT.  
KW DNA-binding; RNA-binding; Transport.  
FT VARIANT 122 F -> L.  
SQ SEQUENCE 265 AA; 29857 MW; 7BE807863924592E CRC64;

Query Match 7.8%; Score 91; DB 1; Length 265;  
Best Local Similarity 24.0%; Pred. No. 2.9;  
Matches 50; Conservative 21; Mismatches 59; Indels 78; Gaps 9;

QY 22 KVKDFLKLS-HEEFLKEEFPKLGKALGIVSVKQSFSEKVRGKIPKIKLKKENFPIQVDYKT 80  
| | | | | : | | | | | : | | | | | : | | : | |  
Db 6 KVSDFLALTKEEILPKALTRKLTVSISTKDVISVKESESLCDIDL--VNVPLD-KYRY 62  
  
QY 81 PETLGTDRVALAYSAKKPYGKNVVISAGTALVIDLVLEGFKGFGFITLGLGKKLKILSD 140  
| | : | | | | | : | | | | | : | | : | |  
Db 63 VGVLG-----VFTGEWLVPDFV-----KGGVTVSVIDKRLE----- 94  
  
QY 141 LAEGIPPEFFPEEVEIFLGRSTRECVLGAGYRESTEFIKSTLKRKVF-----KRKFKV 194  
: : : : : | | | : : : | | : | | : | | : | |  
Db 95 -----NSKECII-GTYRAAAKDRRFQKLVPNYFVSVADAKRK--- 131  
  
QY 195 VITGEGKYFSKFGIYDPLLVHRGMRNL 222  
| | : | | : | | : | | : | | : | | : | | : | |  
Db 132 -----PWQVHVRIQNL 142

RESULT 8  
NAOX\_METJA STANDARD; PRT; 463 AA.  
AC Q58065;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Putative NADH oxidase (EC 1.6.99.3) (NOXase).  
GN MJ0649.  
OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
OC Methanocaldococcaceae; Methanocaldococcus.  
OX NCBI\_TaxID=2190;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
RX MEDLINE=96337999; PubMed=8688087;  
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
RT jannaschii.";  
RL Science 273:1058-1073(1996).  
CC -1- FUNCTION: CATALYZES THE FOUR-ELECTRON REDUCTION OF MOLECULAR  
CC OXYGEN TO WATER (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: 2 NADH + O(2) = 2 NAD(+) + 2 H(2)O.  
CC -1- COFACTOR: FAD.  
CC -1- SIMILARITY: SOME, TO FLAVOPROTEIN OXIDOREDUCTASES. HIGHEST, TO  
CC NADH PEROXIDASE (NPR).  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----



```
DR EMBL; U67512; AAB98641.1; -.
DR TIGR; MJ0649; -.
DR InterPro; IPR001327; FAD_pyr_redox.
DR InterPro; IPR000205; NAD_binding.
DR InterPro; IPR004099; pyr_redox_dim.
DR Pfam; PF00070; pyr_redox; 1.
DR Pfam; PF02852; pyr_redox_dim; 1.
DR ProDom; PD000139; FAD_pyr_redox; 1.
KW Hypothetical protein; Oxidoreductase; NAD; Flavoprotein; FAD;
KW Redox-active center; Complete proteome.
FT ACT_SITE 57 REDOX-ACTIVE.
FT NP_BIND 167 182 NAD (BY SIMILARITY).
FT NP_BIND 286 296 FAD (FLAVIN PART) (BY SIMILARITY).
SQ SEQUENCE 463 AA; 50599 MW; AA216A9F20E6C1AA CRC64;

Query Match 7.8%; Score 91; DB 1; Length 463;
Best Local Similarity 20.5%; Pred. No. 5.4;
Matches 48; Conservative 41; Mismatches 79; Indels 66; Gaps 10;

QY 19 EGKKVKDFLKLSHEE-----FLKEFPKALKAL-----GISVKQSFSEKVRG----- 59
Db 136 EGKDLGDFKVRTIEDGRAILKYIEENGCKKVAVVGAGAIGLEMAVGLKCRGLDLVVEM 195
QY 60 ---KIPKI-----KFLKKENFPIQVDYKTPETLGTDRVALAYSAKKFFGKNNVVI 106
Db 196 APQVLPFRFLDPDMAEIVQKYLEKEGKIVMLSKPLEKIVGKEKVEAVYVDGKLYDVMVIM 255
QY 107 SAGTALVIDLVLEGGKGGFITLGLGKKLKILSDLAEGIPPEFFPEVEIFLGRSTRCEVL 166
Db 256 ATGVRPNIELAKKAGCKIGKFAIEVNEKMQ-----TSIPNIY-----AVGDCV- 298
QY 167 GGAYRESTEFI--KSTLKLWRKVFKRKKVV---ITGGEGKYF-----SKFG 208
Db 299 -----EVIDFITGEKTLSPFGTAARQGVKAGKNIAGVEAKFYPLVNSAVSKIG 347

RESULT 9
YOLL_MOUSE STANDARD; PRT; 379 AA.
AC P11260;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-JUL-1989 (Rel. 11, Last annotation update)
DE Hypothetical protein ORF-1137.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87064284; PubMed=3023821;
RA Loeb D.D., Padgett R.W., Hardies S.C., Shehee W.R., Comer M.B.,
RA Edgell M.H., Hutchison C.A. III;
RT "The sequence of a large l1md element reveals a tandemly repeated 5'
RT end and several features found in retrotransposons.";
RL Mol. Cell. Biol. 6:168-182(1986).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M13002; AAA66023.1; ALT_INIT.
CC PIR; A24906; QQMSLL.
CC InterPro; IPR004244; Transposase_22.
CC Pfam; PF02994; Transposase_22; 1.
KW Hypothetical protein.
SQ SEQUENCE 379 AA; 43788 MW; F4DDE312807644D3 CRC64;

Query Match 7.7%; Score 90.5; DB 1; Length 379;

```

```
Best Local Similarity 23.5%; Pred. No. 4.7;
Matches 51; Conservative 38; Mismatches 73; Indels 55; Gaps 12;

QY 12 SVDIALWEGKKVKDFLKLSHEEFLKEEF--PKLKALGISVKQSFSEKVRKIPKIFLK 69
Db 171 NIDITVKENTKCKRIL-TQNIQVIQDTMRPNRLRIIGIDENEDF--QLKGPANIFNKIIE 227
QY 70 ENFP-----IQVDYKTPETLGTDRVALAYSAKKFFGKNNVVISACTALVIDLVLEG 120
Db 228 ENFPNKKEMPMIIEAYRTPNRLDQKRNSRH-----IIRTNALNKDRILKA 277
QY 121 KFKGGFITLGLGKKLKILSDLAEGIPPEFFPEVEIFLGRSTRCEVLGGAYRESTEFIKST 180
Db 278 VREKGVQTY-KGRPIRI-----TPDFSPETMK-----ARRAWTDVIO-T 314

181 IKLW----RKVFKRKVKVITGGEGKYF---SKFGIY 210
315 LREHKCQPRLLYPAKLSITI-DGETKVFHDKTKFTQY 350

RESULT 10
DCUP_SCHPO STANDARD; PRT; 370 AA.
AC Q9USJ5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Uroporphyrinogen decarboxylase (EC 4.1.1.37) (URO-D) (UPD).
GN SPCC4B3.05C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=972;
RC MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Welltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McComb W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -!- CATALYTIC ACTIVITY: Uroporphyrinogen-III = coproporphyrinogen + 4
CC CO(2).
CC -!- PATHWAY: PORPHYRIN AND HEME BIOSYNTHESIS.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE UROPORPHYRINOGEN DECARBOXYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```





Db 128 KKLSK-----DMLTKEVLGEVTRDPDGLYTTLVVDQYERCLGLVYSSKKSIK---A 177

QY 106 ISAGTALVI-----DLVLEGGKFGGFTLGLGKK-LKILSDLAEGIPEFFPEVEI-FLG 158

Db 178 IDLGRGVYYSRSRNEIWKGE-----TSGNGQKLLQISTDCDSDALKFIVEQENVGFCH 231

QY 159 RSTRECVLGGAYRESTEFIKSTLK 182

Db 232 LETMSCF--GEFKHGLVGLSLLK 253

RESULT 12

SYFA\_METJA

ID SYFA\_METJA STANDARD; PRT; 480 AA.

AC Q57911;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Phenylalanyl-tRNA synthetase alpha chain (EC 6.1.1.20) (Phenylalanine-tRNA ligase alpha chain) (PHERS).

DE -tRNA ligase alpha chain) (PHERS).

3N PHES OR MJ0487.

OS Methanococcus jannaschii.

OC Archaea; Euryarchaeota; Methanococci; Methanococcales;

OC Methanocaldococcaceae; Methanocaldococcus.

OX NCBI\_TaxID=2190;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;

RX MEDLINE=96337999; PubMed=8688087;

RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D., Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D., Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I., Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A., Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D., Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C., Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M., Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.; "Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.";

RL Science 273:1058-1073(1996).

CC -!- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP + diphosphate + L-phenylalanyl-tRNA(Phe).

CC -!- SUBUNIT: Tetramer of two alpha and two beta chains (by SIMILARITY).

CC -!- SUBCELLULAR LOCATION: Cytoplasmic.

CC -!- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY. PHE-TRNA SYNTHETASE ALPHA CHAIN SUBFAMILY 2.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; U67499; AAB98478.1; -.

DR HSSP; P27001; 1PYS.

DR TIGR; MJ0487; -.

DR InterPro; IPR002106; AaTrna\_ligaseII.

DR InterPro; IPR004529; PheS.

DR InterPro; IPR002319; tRNA-synt\_2d.

DR Pfam; PF01409; tRNA-synt\_2d; 1.

DR TIGRFAMs; TIGR00468; pheS; 1.

DR PROSITE; PS50862; AA\_TRNA\_LIGASE\_II; 1.

KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding; Complete proteome.

KW Complete proteome.

SQ SEQUENCE 480 AA; 57027 MW; 1DC2599CF0D97B72 CRC64;

Query Match 7.6%; Score 88.5; DB 1; Length 480;

Best Local Similarity 20.9%; Pred. No. 8.8;

Matches 49; Conservative 42; Mismatches 75; Indels 69; Gaps 11;

QY 13 VDIALWEGKKVKDFLKLKLSHEEF-----LKEEFPKLKALGIS--VKQSFSEKVRGKIPK-IK 65

Db 1 MELHIDEKRLKIFQDNNRDEFNLNELEKFMPEKILRVSWLKGNLVETEKKVKKI 60

QY 66 FLKKENFP-----IQVDYKT-PETLGTDRVALAYSAKKFYGNVVISAGTA 111

Db 61 LIKEEFPERKIANYLKQHNKEIEIKNLKDILPKEEINAALGAIK--RKGIARIEKGKI 118

QY 112 L-----VIDLVLEKFE-----KGGFITLGLGKKLKILSDLA 143

Db 119 IFDNLDYKDVVEQLQKIKENKYLDLDFSEEKKIIDILKRGYVDFDEKEIKI--KLTE 176

QY 144 GIPEFF--PEEVEIFLGRSTRCEVLGGAYRESTEFIKSTLKLWRKVKFKFVVI 196

Db 177 KGKEFIKNPIEIEEITQLTRDIISGK-----WKKAYIRPYDVKV 217

RESULT 13

RS6E\_SULTO

ID RS6E\_SULTO STANDARD; PRT; 213 AA.

AC Q975N7;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE 30S ribosomal protein S6e.

GN RPS6E OR ST0382.

OS Sulfolobus tokodaii.

OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;

OC Sulfolobus.

OX NCBI\_TaxID=111955;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=JCM 10545 / 7;

RX MEDLINE=21456156; PubMed=11572479;

RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S., Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y., Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A., Oshima T., Kikuchi H.;

RT "Complete genome sequence of an aerobic thermoacidophilic Crenarchaeon, Sulfolobus tokodaii strain 7.";

RL DNA Res. 8:123-140(2001).

CC -!- SIMILARITY: BELONGS TO THE S6E FAMILY OF RIBOSOMAL PROTEINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AP000982; BAB65363.1; ALT INIT.

DR InterPro; IPR001377; Ribosomal\_S6E.

DR Pfam; PF01092; Ribosomal\_S6e; 1.

DR ProDom; PD003460; Ribosomal\_S6E; 1.

DR PROSITE; PS00578; RIBOSOMAL\_S6E; 1.

KW Ribosomal protein; Complete proteome.

SQ SEQUENCE 213 AA; 23636 MW; 90FA15FB3E3F36FF CRC64;

Query Match 7.5%; Score 88; DB 1; Length 213;

Best Local Similarity 26.3%; Pred. No. 3.8;

Matches 46; Conservative 20; Mismatches 43; Indels 66; Gaps 10;

QY 19 EGKKVKDFLKLKLSHEEFKKEEFPKLKALGI-----SVKQSFSEKVRGKIPKIFLKKN 71

Db 37 EGKAL-PIAKINQK--LKEE-----LGLDTLLTLQTIKQEGDKVK-----VKTH 78

QY 72 FPIQVDYKTPETLGTDRVALAYSAKKFYGNV-----VISAGT 110

Db 79 FKIEDNNVP-----SDEVWIAASIAEKYGANEFEAEAYRTKSFQSLDSQKLSNIIGAKI 134







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:36:04 ; Search time 8.35258 Seconds  
(without alignments)  
2635.685 Million cell updates/sec

Title: US-09-813-453A-12  
Perfect score: 1168  
Sequence: 1 MRFLTVDVGNSSVDIALWEG.....YDPLLVRGMRNLLLYLHRI 229

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1168	100.0	229	2 E70465	hypothetical prote
2	193.5	16.6	246	2 D72320	conserved hypothet
3	177.5	15.2	209	2 H81382	hypothetical prote
4	156.5	13.4	259	2 AF1102	conserved hypothet
5	149.5	12.8	259	2 AF1464	conserved hypothet
6	141.5	12.1	262	2 F70165	conserved hypothet
7	138.5	11.9	223	2 F64627	hypothetical prote
8	138.5	11.9	254	2 F83660	hypothetical prote
9	138.5	11.9	273	2 E97293	probable transcrip
10	133.5	11.4	262	2 E75516	conserved hypothet
11	133	11.4	257	2 S75559	hypothetical prote
12	131.5	11.3	223	2 G71887	hypothetical prote
13	129.5	11.1	261	2 B87489	transcription acti
14	128.5	11.0	265	2 T36391	hypothetical prote
15	122.5	10.5	273	2 D71326	conserved hypothet
16	117	10.0	233	2 S66100	conserved hypothet
17	117	10.0	276	2 AI2292	hypothetical prote
18	114	9.8	272	2 A70955	hypothetical prote
19	112	9.6	274	2 H86937	conserved hypothet
20	108	9.2	592	2 H82031	probable blotin-la
21	105	9.0	224	2 A99571	conserved hypothet
22	103	8.8	592	2 B81009	BirA protein/Bvg a
23	102.5	8.8	254	2 G69495	transcription regu
24	101	8.6	596	2 D84972	ABC transporter AT
25	100.5	8.6	733	1 B30001	ribosomal protein
26	98.5	8.4	248	2 H83111	hypothetical prote
27	98	8.4	390	2 A53773	La/SS-B homolog D-
28	98	8.4	491	2 F64118	cytosolic axial fi
29	97.5	8.3	484	2 F71954	guanosine-5'-triph

30	96	8.2	390	2 A53781	ribonucleoprotein
31	96	8.2	578	2 C97167	uncharacterized co
32	95.5	8.2	426	1 B71249	hypothetical prote
33	95	8.1	242	2 A82637	conserved hypothet
34	94.5	8.1	484	2 F64554	guanosine pentapho
35	93.5	8.0	427	2 T39113	probable flavohemo
36	92	7.9	194	2 F71836	anthranilate synth
37	91	7.8	265	2 S65054	movement protein -
38	91	7.8	463	2 A64381	NADH oxidase - Met
39	90.5	7.7	235	2 H81451	ubiquinone/menaqui
40	90.5	7.7	357	1 QQMSLL	retrovirus-related
41	90	7.7	370	2 T50443	probable uroporphy
42	89	7.6	286	2 B70430	transcription regu
43	89	7.6	799	1 SHBY	phosphoribosyl-AMP
44	88.5	7.6	462	2 D85438	hypothetical prote
45	88.5	7.6	470	2 T45962	hypothetical prote

ALIGNMENTS

RESULT 1

E70465

hypothetical protein aq\_1924 - Aquifex aeolicus

C:Species: Aquifex aeolicus

C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 05-Nov-1999

C:Accession: E70465

R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666; PMID:9537320

A:Accession: E70465

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-229 <AQF>

A:Cross-references: GB:AE000763; NID:g2984178; PIDN:AAC07720.1; PID:g2984188; GB:AE

A:Experimental source: strain VF5

C:Genetics:

A:Gene: aq\_1924

Query Match 100.0%; Score 1168; DB 2; Length 229;  
Best Local Similarity 100.0%; Pred. No. 2.6e-82;  
Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRFLTVDVGNSSVDIALWEGKKVKDFLKLSHEEFLKEEPPKALKALGISVKQSEKVRGK 60

Db 1 MRFLTVDVGNSSVDIALWEGKKVKDFLKLSHEEFLKEEPPKALKALGISVKQSEKVRGK 60

QY 61 IPKIKFLKKENFPIQVDYKTPETLGTDRVALAYSAKKFYGNVVISAGTALVIDLVLEG 120

Db 61 IPKIKFLKKENFPIQVDYKTPETLGTDRVALAYSAKKFYGNVVISAGTALVIDLVLEG 120

QY 121 KFKGGFITLGLGKKLKILSDLAEGIPPEFFPEVEIFLGRSTRECVLGGAYRESTEFIKST 180

Db 121 KFKGGFITLGLGKKLKILSDLAEGIPPEFFPEVEIFLGRSTRECVLGGAYRESTEFIKST 180

QY 181 LKLWRKVFRKFKVVTGEGKYFSKFGIYDPLLVRGMRNLLLYLHRI 229

Db 181 LKLWRKVFRKFKVVTGEGKYFSKFGIYDPLLVRGMRNLLLYLHRI 229

RESULT 2

D72320

conserved hypothetical protein - Thermotoga maritima (strain.MSB8)

C:Species: Thermotoga maritima

C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000

C:Accession: D72320

R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; H  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson  
C.M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genom



F64627

F64627

QY 101. KVVVISAGTALVIDLVLECK - FKGGFTTGLGKKLILSDLAEGIPFFFF

QY 101. KNTVVISAGTALVIDLVLECK-FKGGFITLGLGKKLKILSDLAEGIPPEFFPEEVEIFLGR 159



Db 123 YPAIVDFGTATTYCLINEKKQYAGGVIAPGIMISTEALYHRASKLPRIEIAKPKQVVGT 182

QY 160 STRECVLGGAYRESTEFIKSTLKLWRKVKFKKVVITGGEGKYF---SKFGIYDPLLV 215

Db 183 NTIDSMQSGIFYGYVSQVDGVVKRMKAQAESEPKVIATGGLAKLIGTESETIDVIDSFLT 242

QY 216 HRGMRNLLY 224

Db 243 LKGLQ-LIY 250

RESULT 9

E97293

probable transcription regulator, homolog of Bvg accessory factor [imported] - Clostridi

C;Species: Clostridium acetobutylicum

C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001

C;Accession: E97293

R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,

R;Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo

A;Reference number: A96900; MUID:21359325; PMID:21359325

A;Accession: E97293

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-273 <KUR>

A;Cross-references: GB:AE001437; PIDN:AAK81136.1; PID:gl5026270; GSPDB:GN00168

A;Experimental source: Clostridium acetobutylicum ATCC824

C;Genetics:

A;Gene: CAC3200

C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 11.9%; Score 138.5; DB 2; Length 273;

Best Local Similarity 21.6%; Pred. No. 0.0021;

Matches 58; Conservative 54; Mismatches 103; Indels 53; Gaps 11;

QY 4 LTVDVGNSSVDIALWEGKKVKVDFLKLSHEEFLKEEFPKLGISVKQSFSEK----- 56

Db 14 LVLVDGNTNIVLGIYNDTKLTAEWRLSTDVLRSD-----EYGIQVMNLFQQDKLDPTLV 68

QY 57 -----VRGKIPKIKF---LKKNFP-----IQVDYKTPETLGTDRVALAYSA 95

Db 69 EGVLISSVVPNIMYSLEHMIRKYFKINPLVVGPGIKTGINIKYDNPKVEGADRIVNAVA 128

QY 96 KKFYGNVVISAGTALVIDLVLE-GKFKGGFITLGLGKKLKILSDLAEGIPPEFFPEEVE 154

Db 129 HEYKRSLSIIIDFGTATTCFAVRENGDYLGAICPGIKVSSEALFEKAALP-----RVE 183

QY 155 IF-----LGRSTRECVLGG---AYRESTEFIKSTLK-LWRKVFKRKFKVVITGGEGKYFS 205

Db 184 LKPAYAICKNTISSIQSGIVGYIGQVRYIVERMKEELQEEGEKEPLVVATGGLAKLIS 243

QY 206 K----FGIYDPLLVHRGMRNLLYLYHRI 229

Db 244 EEAKNVDVINPFLTLEGLR-LIYEKNRV 270

RESULT 10

E75516

conserved hypothetical protein - Deinococcus radiodurans (strain R1)

C;Species: Deinococcus radiodurans

C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 20-Jun-2000

C;Accession: E75516

R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A;Reference number: A75250; MUID:20036896; PMID:10567266

A;Accession: E75516

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-262 <WHI>

A;Cross-references: GB:AE001905; GB:AE000513; NID:g6458144; PIDN:AAF10040.1; PID:g64

A;Experimental source: strain R1

C;Genetics:

A;Gene: DR0461

A;Map position: 1

C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 11.4%; Score 133.5; DB 2; Length 262;

Best Local Similarity 23.0%; Pred. No. 0.0049;

Matches 58; Conservative 44; Mismatches 113; Indels 37; Gaps 9;

QY 4 LTVDVGNSSVDIALWE--GKKVKDFLKLSSHEEFLKEE-----FPK---LK 43

Db 7 LAVDIGNTTVLGLADASGALTHTWRIRTNREMLPDDLALQLHGLFTLAGAPIPRAAVLS 66

QY 44 ALGISVKQSFSEKVRGKIPKIKF-LKKENFP-IQVDYKTPETLGTDRVALAYSAKKFGY- 100

Db 67 SVAPPVGENYALALKRHFMDIDAFAVSAENLDPDVTVELDTPGSVGADRLCNLFGAEKYLG 126

QY 101 -KNVVVISAGTALVIDLVLE-GKFKGGFITLGLGKKLKILSDLAEGIPPEFFPEEVEIFLG 158

Db 127 LDYAVVDFGTSTNFDVVGRRRFLGGILATGAQVSADALFARAALKPRITLQAPETAIG 186

QY 159 RSTRECVLGGAYRESTEFIKSTLKLWRKVFKRKFKVVITGGEGKYFSK-----FGIY 210

Db 187 KNTVHALQSGLVFGYAEVMDGLLRIRAEPLPGEAVAVATGG----FSRTVQICQEIY 242

QY 211 DPLLVHRGMRNL 222

Db 243 DEFTLRLGLVEL 254

RESULT 11

S75559

hypothetical protein slr0812 - Synechocystis sp. (strain PCC 6803)

C;Species: Synechocystis sp.

A;Variety: PCC 6803

C;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999

C;Accession: S75559

R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,

O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Ya

DNA Res. 3, 109-136, 1996

A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocy

S.

A;Reference number: S74322; MUID:97061201; PMID:8905231

A;Accession: S75559

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-257 <KAN>

A;Cross-references: EMBL:D90911; GB:AB001339; NID:g1653083; PIDN:BAA18120.1; PID:d1c

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C;Genetics:

A;Start codon: GTG

Query Match 11.4%; Score 133; DB 2; Length 257;

Best Local Similarity 28.3%; Pred. No. 0.0052;

Matches 66; Conservative 31; Mismatches 118; Indels 18; Gaps 8;

QY 3 FLTVDVGNSSVDIALWEGKK-----VKDFLKLSSHEEFLKEEFPKLGISVKQSFSEKV 57

Db 20 WLGLMIGNSRHLWAYCSGNAPLQTVWTDYNPKSAQ--LPVLVGKVPPLMLASVWPEQTEVW 77

QY 58 RGKIPKIKFLKKENFP IQVDYKTPETLGTDRVALAYSAKKFGYKNVVVISAGTALVI-DL 116

Db 78 RYQPKILTILK--NLPLVNLV---PSFGIDRALAGLGTGLTYGFPCLVVDGGTALTITGF 132

QY 117 VLEGKFKGGFITLGLGKKLKILSDLAEGIPPEFFPEEVEIFLGR---STRECVLGGAYRES 173

Db 133 DQDKLVGGAILPLGLQLATLGLDRLAALPKLEMDQLTELDPDRWALDTPSAIFSGVVYGV 192

QY 174 TEFIKSTLKLWRKVFKRKFKVVITGGEGKYFSKF-GIYDPLLVHRGMRNLLYL 225

Db 193 LGALQSYLQDWQKLFPGA-AMVITGGDGKILHGFLKEHSPNLSVAWDDNLIFL 244



RESULT 12  
G71887  
hypothetical protein jhp0796 - Helicobacter pylori (strain J99)  
C:Species: Helicobacter pylori  
A:Variety: strain J99  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 18-Aug-2000  
C:Accession: G71887  
R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;  
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;  
Nature 397, 176-180, 1999  
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path  
A:Reference number: A71800; MUID:99120557; PMID:9923682  
A:Accession: G71887  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-223 <ARN>  
A:Cross-references: GB:AE001509; GB:AE001439; NID:g4155350; PIDN:AAD06372.1; PID:g415536  
;Experimental source: strain J99  
;Genetics:  
A:Gene: jhp0796  
C:Superfamily: Campylobacter jejuni hypothetical protein Cj0394c

Query Match 11.3%; Score 131.5; DB 2; Length 223;  
Best Local Similarity 26.5%; Pred. No. 0.0057;  
Matches 61; Conservative 32; Mismatches 84; Indels 53; Gaps 11;  
QY 7 DVGNSVDIALWEGKKVKDFLKLSSHEEFLKEEFPKALKG-----ISVKQSFSEKVRG 59  
Db DIGNTRIIFA-----QNYQLFSSAKEDLKLGIQKEIFYISVNEENEKALLN 63  
QY 60 KIPKIKFLKKNFPIQVDYKTPETLGTDRVALAYSAAKPYGKNVVISAGTALVIDLVLE 119  
Db CYPNAKNI-AGFFHLETDY---IGLGIDROMACLAIV-----VNGVIVDAGSAITIDLVE 114  
QY 120 GKFKGGFITLGLG-----KKLKILSDLAEGIPPEFFPEVEIFLGRSTRECVLGGAYRE 172  
Db GKHLGGCILPGLAQYVHAYKSAKILEQPFKAL-----DSLEV-LPKNTRDAVNYG----- 164  
QY 173 STEFIKSTLKLWRKVKFRKPKVVTGEGGKYFSKF---GIYDPLLVRHGM 219  
Db MILSIISCIQHLAKDQ-KIYLCGGDAKYLSAFLPHSVCKERLVFDGM 210

RESULT 13  
B87489  
transcription activator, probable Baf family [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
C:Accession: B87489  
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: B87489  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-261 <STO>  
A:Cross-references: GB:AE005673; NID:g13423392; PIDN:AAK23910.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CC1935  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 11.1%; Score 129.5; DB 2; Length 261;  
Best Local Similarity 22.1%; Pred. No. 0.0098;  
Matches 57; Conservative 40; Mismatches 128; Indels 33; Gaps 8;  
QY 1 MRPLTVDVGNSSVDIALWEGKK-VKDFLKLSSHEEFLKEEFP-----KLKALGISVKQS 52  
Db MMLLAIEQGNNTMTFAIHDGASVVAQWRSATESTRTADEYVWLSQLLSMQGLGFRAIDA 60

QY 53 --FSEKVRGKIPKIKFLKKNFPIQ-----VDYKTPETLGTDRVALAYSAKK 97  
Db VIISVVVQSFIFNLRLNSRRYFNVFVLVIGENAKLGIDVRIEKPSEACADRLVNAIGAAM 120  
QY 98 FYGKNVVVISAGTALVIDLV-LEGKFKGGFITLGLGKKLKLSDLAEGIPPEFFPEEV--E 154  
Db VYPGPLVVIDSGTATTDFDIVAADGAFEGGIIAPGINLSMQALHEAAKLPRIAIORPAGN 180  
QY 155 IFLGRSTRECVLGGAYRESTEFIKSTLKLWRKVKFRKPKVVTGEGKYP-----SKFGIY 210  
Db RIVGTDTVSAMQSGVFWGYISLIEGLVARIKAEGERPMTVIATGGVASLFEGATDSIDHF 240  
QY 211 DPLLVRHGRMNRNLLYLYHR 228  
Db DSDLTIRG---LLEIYRR 255

RESULT 14  
T36391  
hypothetical protein SCE94.31c - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jun-2000  
C:Accession: T36391  
R;Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M  
submitted to the EMBL Data Library, April 1999  
A:Reference number: Z21573  
A:Accession: T36391  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-265 <OLI>  
A:Cross-references: EMBL:AL049628; PIDN:CAB40880.1; GSPDB:GN00070; SCOEDB:SCE94.31c  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SCE94.31c  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 11.0%; Score 128.5; DB 2; Length 265;  
Best Local Similarity 24.1%; Pred. No. 0.012;  
Matches 49; Conservative 36; Mismatches 71; Indels 47; Gaps 9;  
QY 4 LTVDVGNSSVDIALWEGKKVKDFLKLSSHEEFLKEEFPKALK----- 45  
Db LTIDVGNTHTVLGLFDGEDIVEHWRISTDS--RTADELAVLLQGLMGHPLLGLDELGDG 60  
QY 46 --GI-----SVKQSFSEKVR---GKIPKIKFLK--KENFPQVDYKTPETLGTDRVAL 91  
Db IDGIAICATVPSPVLHRELRETVRRYYGDVPAVLVEPVGKTVGPILTDH--PKEVGADRIIN 118  
QY 92 AYSAKPKYGNVWVISAGTALVIDLV-LEGKFKGGFITLGLGKKLKLSDLAEGIPPEFFP 150  
Db AVAAVELYGGPAIVDFGTATTDFDAVSARGEYIGGVIAPGIEISVEAL-----GVKGAQL 173  
QY 151 EEVEI-----FLGRSTRECVLGG 168  
Db RKIEVARPRSVIGKNTVEAMQSG 196

RESULT 15  
D71326  
conserved hypothetical protein TP0431 - syphilis spirochete  
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)  
C:Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: D71326  
R;Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.;  
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.;  
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
Science 281, 375-388, 1998  
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.  
A:Reference number: A71250; MUID:98332770; PMID:9665876  
A:Accession: D71326  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA

Query Match	10.5%;	Score 122.5;	DB 2;	Length 273;
Best Local Similarity	23.7%;	Pred. No. 0.035;		
Matches 62;	Conservative 48;	Mismatches 105;	Indels 47;	Gaps 12;

Search completed: June 24, 2003, 22:03:11  
Job time : 10.3526 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:26 ; Search time 23.511 Seconds  
(without alignments)  
1297.879 Million cell updates/sec

Title: US-09-813-453A-12  
Perfect score: 1168  
Sequence: 1 MRFLTVDVGNSSVDIALWEG.....YDLLVHRGMRNLLYLHRI 229

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues  
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002:\*  
1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1168	100.0	229	23	AAU911159	Aquifex aeolicus p
2	225	19.3	257	23	AAU911174	Pantothenate kinas
3	193.5	16.6	246	23	AAU911156	Thermotoga maritim
4	177.5	15.2	209	23	AAU911165	Pantothenate kinas
5	167.5	14.3	262	23	AAU911170	Pantothenate kinas
6	164.5	14.1	241	23	AAU911179	Pantothenate kinas
7	158.5	13.6	258	23	AAU911153	Rhodobacter capsul
8	156.5	13.4	259	23	ABB47661	Listeria monocytog
9	154.5	13.2	258	22	AAU01243	B. subtilis novel
10	154.5	13.2	258	23	AAU911149	Bacillus subtilis

11	148.5	12.7	258	23	AAU911172	Pantothenate kinas
12	141.5	12.1	262	23	AAU911158	Borrelia burgdorfe
13	139.5	11.9	255	23	AAU911154	Geobacter sulfurre
14	138.5	11.9	223	19	AAW98422	H. pylori GHPO 344
15	138.5	11.9	223	23	AAU911161	Helicobacter pylor
16	138.5	11.9	223	23	AAU911181	Pantothenate kinas
17	138.5	11.9	254	23	AAU911171	Pantothenate kinas
18	133.5	11.4	262	23	AAU911155	Deinococcus radiop
19	133	11.4	256	23	AAU911175	Pantothenate kinas
20	133	11.4	257	23	AAU911160	Synechocystis pant
21	132.5	11.3	244	23	AAU911168	Pantothenate kinas
22	130	11.1	219	23	AAU911176	Pantothenate kinas
23	128.5	11.0	265	23	AAU911151	Streptomyces coeli
24	125.5	10.7	260	23	AAU911173	Pantothenate kinas
25	123.5	10.6	250	23	AAU911150	Clostridium acetob
26	122.5	10.5	273	23	AAU911157	Treponema pallidum
27	120.5	10.3	212	23	AAU911177	Pantothenate kinas
28	117	10.0	233	23	AAU911163	Pantothenate kinas
29	114	9.8	272	22	AAG81225	Mycobacterium tube
30	114	9.8	272	23	AAU911152	Mycobacterium tube
31	111.5	9.5	389	21	AAU74909	Neisseria meningit
32	108	9.2	455	21	AAU74910	Neisseria meningit
33	108	9.2	592	20	AAU38616	Neisseria meningit
34	108	9.2	592	21	AAU74913	Neisseria meningit
35	108	9.2	592	23	AAU911166	Pantothenate kinas
36	103	8.8	592	20	AAU38615	Neisseria meningit
37	103	8.8	592	21	AAU74912	Neisseria meningit
38	103	8.8	592	23	AAU911169	Pantothenate kinas
39	99	8.5	455	20	AAU38617	Neisseria gonorrh
40	99	8.5	455	21	AAU74908	Neisseria gonorrh
41	99	8.5	460	23	AAU911167	Pantothenate kinas
42	99	8.5	592	20	AAU38618	Neisseria gonorrh
43	99	8.5	592	21	AAU74911	Neisseria gonorrh
44	98.5	8.4	248	23	AAU911164	Pantothenate kinas
45	97.5	8.3	484	18	AAW55720	H. pylori ORF 07ee

ALIGNMENTS

RESULT 1  
AAU911159  
ID AAU911159 standard; Protein; 229 AA.  
XX  
AC AAU911159;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Aquifex aeolicus pantothenate kinase Coax.  
XX  
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX  
OS Aquifex aeolicus.  
XX  
PN WO200216601-A2.  
XX  
PD 28-FEB-2002.  
XX  
PF 24-AUG-2001; 2001WO-US26531.  
XX  
PR 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX  
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX  
PI Yocum RR, Patterson TA;  
XX  
DR WPI; 2002-269358/31.  
XX  
PT Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -

Claim 10; Page 77; 128pp; English.

The invention describes assays for identifying a (potential) antibiotic comprising contacting an assay composition comprising a pantothenate kinase (Coax) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulating agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (Coax) protein described in the invention.

RESULT 2	
AAU91174	AAU91174 standard; Protein; 257 AA.
ID	
XX	
AC	
XX	AAU91174;
DT	05-JUN-2002 (first entry)
XX	
DE	Pantothenate kinase (Coax) #12.
XX	
KW	Pantothenate kinase; Coax; antiblotic; antimicrobial;
WZ	pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX	Chlorobium tepidum.
PN	WO200216601-A2.
XX	
PD	28-FEB-2002.
XX	
PF	24-AUG-2001; 2001WO-US26531.
XX	
PR	24-AUG-2000; 2000US-227860P.
PR	20-MAR-2001; 2001US-0813453.
XX	
PA	(OMNI-) OMNIGENE BIOPRODUCTS INC.
XX	
PI	Yocum RR, Patterson TA;
XX	
DR	WPI; 2002-269358/31.
DR	N-PSDB; ABK54195.
XX	
PT	Identifying potential antibiotic or antimicrobial agent, comprises
PT	contacting composition comprising pantothenate kinase (Coax) protein
PT	with test compound and identifying inhibitor of the Coax protein -
XX	
PS	Claim 10; Page 103-104; 128pp; English.

	The invention describes assays for identifying a (potential) antibiotic comprising contacting an assay composition comprising a pantothenate kinase (Coax) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulating agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (Coax) protein described in the invention.
CC	
XX	
SQ	Sequence    257 AA;
	Query Match                  19.3%; Score 225; DB 23; Length 257; Best Local Similarity       27.7%; Pred. No. 1.2e-14; Matches 71; Conservative    39; Mismatches 104; Indels 42; Gaps 8;
QY	4 LTVDVGNSSVDIALWEGKKVKDFLKLSEEF-----LKKEF----- 39     :     :   :   :   :    :
Dd	3 LVVDIGNTSTTLAFTGDEEPSVESVPALFADSSSTMREVFENMARKHGEPQAIAICSV 62   L V D I G N T S T T L A F T G D E E P S V E S V P A L F A D S S T M R E V F E N M A R K H G E P Q A I A I C S V
QY	40 PKLKALGISVKQS-FSEKVRGKIPIKFLKKENFIQVDYKTPTLTGTDRVAL-AYSAKK 97    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Dd	63 PSATAVGSALESIFS-----VPVLTCCKLRFPFRLDYATPHTFGADRLLCAWSRHL 116    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY	98 FYGKNVVVISAGTALVIDLV-LEGKFKGGFITGLGKKILKILSDLAEGIPFEFFVEEIF 156       :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Dd	117 FSEKPVIADVDIGTAIFDVLDTVGNRYRGGLIMPGDIMMAGALHSRTAQLPQVRIDRPESL 176         :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY	157 LGRSTRECVLGAAYRESTEFIKSTLKLWRKVFKRK-----KVVITGGEGKY----FSKF 207           :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Dd	177 LGRSTTECIKSGVFWGVVKQIGGLVDAIRGDLVRDFGESTVEIVTGGSRIIIVPEIGPV 236           :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY	208 GIYPDPLLVHGRMRNLL 223 :
Dd	237 SVIDEAVLGSRDLLL 252 :

RESULT 3	
AAU91156	
ID	AAU91156 standard; Protein; 246 AA.
XX	
AC	AAU91156;
XX	
DT	05-JUN-2002 (first entry)
XX	
DE	Thermotoga maritima pantothenate kinase CoaX.
XX	
KW	Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW	pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX	
OS	Thermotoga maritima.
XX	
PN	WO200216601-A2.
XX	
PD	28-FEB-2002.
XX	
PF	24-AUG-2001; 2001WO-US26531.
XX	
PR	24-AUG-2000; 2000US-227860P.
PR	20-MAR-2001; 2001US-0813453.
XX	
PA	(OMNI-) OMNIGENE BIOPRODUCTS INC.
XX	
PI	Yocum RR, Patterson TA;
XX	
DR	WPI; 2002-269358/31.
XX	
PT	Identifying potential antibiotic or antimicrobial agent, comprises
PT	contacting composition comprising pantothenate kinase (CoaX) protein
PT	with test compound and identifying inhibitor of the CoaX protein -
XX	



PS Claim 10; Page 74-75; 128pp; English.  
XX  
CC The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.  
XX  
SQ Sequence 246 AA;

Query Match 16.6%; Score 193.5; DB 23; Length 246;  
Best Local Similarity 26.5%; Pred. NO. 1.8e-11;  
Matches 66; Conservative 45; Mismatches 105; Indels 33; Gaps 6;  
/ 3 FLTVDVGNSSVDIALWEGKKVKDFLKLSSHEEFLKEE-----FPKLKALGI-- 47  
Db 2 YLLVDVGNTHSVFSTEDGKTFRRWRLSTGVFQTEDELFSLHPLLLGDAMREIKGIGVAS 61  
QY 48 -----SVKQSFSEKVRGKIPKIKFLKKENFPIQVDYKTPETLGTDRVALAYSAKKFGK 101  
Db 62 VVPTQNTVIERFSQKFFHISP--IWKAKNGCVKWNKPNSEVGADRVANVAVFVKEYGK 119  
QY 102 NVVVISAGTALVIDLVLEGKFKGGFITLGLGKKLKILSDLAEGIPFPEEVEIFLGRST 161  
Db 120 NGIIDMGATTVDLVVNGSYEGGAILPGFFMMVHSLFRGTAKLPLVEVKPADVFWGKDT 179  
QY 162 RECVLGAYRESTEFIKSTLKLWRKVKFRKFKVVITGGEGKYFS---KFGIYDPLLVRHG 218  
Db 180 EENIRLGVVNGSVYALEGIIIGRIKEY-GDLPVVLTTGGQSKIVKDMIKHEIFDEDLTIKG 238  
QY 219 MRNLLYLYH 227  
Db 239 -----VYH 241

RESULT 4  
AAU911165  
ID AAU911165 standard; Protein; 209 AA.  
XX  
AC AAU911165;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Pantothenate kinase (Coax) #3.  
XX  
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX  
OS Campylobacter jejuni.  
XX  
PN WO200216601-A2.  
XX  
PD 28-FEB-2002.  
XX  
PF 24-AUG-2001; 2001WO-US26531.  
XX  
PR 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX  
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX  
PI Yocum RR, Patterson TA;  
XX  
DR WPI; 2002-269358/31.  
XX  
PT Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -

XX Claim 6; Page 84-85; 128pp; English.  
XX  
CC The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.  
XX  
SQ Sequence 209 AA;

Query Match 15.2%; Score 177.5; DB 23; Length 209;  
Best Local Similarity 25.8%; Pred. No. 6.3e-10;  
Matches 62; Conservative 44; Mismatches 83; Indels 51; Gaps 10;  
QY 4 LTVDVGNSSVDIALWEGK----KVKDFLKLSSHEEFLKEEFPKLKALGISVKQSFSEKVRG 59  
Db 2 LLCDIGNSNANF-LDDNKYFTLNIDQFLEFKNEQ-----KIFYINVNEHLKEHLK- 50  
QY 60 KIPKIKFLKKENFPIQVDYKTPET----LGTDRVALAYSAKKFKYGNVNVVISAGTALVID 115  
Db 51 -----NQKNFINLEPYFLEFDITYQGLGIDRIAACYTI-----EDGVVVDAGSAITID 97  
QY 116 LVLEGKFKGGFITLGLGKKLKILSDLAEGIPFPEEVEI-FLGRSTRECVLGGAYREST 174  
Db 98 IISNSIHLGGFILPGIANYKKIYSHISPRKSEFNTQVSLDAFPQKMDALSYGVFKGIV 157  
QY 175 EFTKSTLKLWRKVKFRKFKVVITGGEGKYFSKF--GIYDPLLVRHGMR-----NLLY 224  
Db 158 LLIKDA-----AQNKKLYFTGGDQGFANYFDHAIYDKLLIFRGMKKIIKENPNLLY 209

RESULT 5  
AAU911170  
ID AAU911170 standard; Protein; 262 AA.  
XX  
AC AAU911170;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Pantothenate kinase (Coax) #8.  
XX  
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX  
OS Bacillus anthracis.  
XX  
PN WO200216601-A2.  
XX  
PD 28-FEB-2002.  
XX  
PF 24-AUG-2001; 2001WO-US26531.  
XX  
PR 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX  
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX  
PI Yocum RR, Patterson TA;  
XX  
DR WPI; 2002-269358/31.  
DR N-PSDB; ABK54191.  
XX  
PT Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -  
XX  
PS Claim 8; Page 98-99; 128pp; English.





[illegible][illegible]



```
XX 28-FEB-2002.
PD
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
DR N-PSDB; ABK54193.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein -
XX
S Claim 10; Page 101-102; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX
SQ Sequence 258 AA;
Query Match 12.7%; Score 148.5; DB 23; Length 258;
Best Local Similarity 23.9%; Pred. No. 7.9e-07;
Matches 60; Conservative 43; Mismatches 111; Indels 37; Gaps 9;
QY 6 VDVGSSVDIALWEGKKVKDF--LKLSHEEFLEKEFPKALKALGTSVKQSFSEKVRG---- 59
Db 5 LDVGNTNTVLGYDGDDELKHHWRIETSRSKTEDEYGMKALKLNHVGLQFSD-IRGIIS 63
QY 60 -KIPKIKFLKE-----NFPIQVDYKTPETLGTDRVALAYSAKKFFYK 101
Db 64 SVVPPINFALERMLCKYFHFKPLIVGPGIKTGLDIKYDNPREVGADRIVNAVAGIHLGYS 123
QY 102 NVVVISAGTALVIDLVLEK-FKGGFITLGLGKKLKILSDLAEGIPPEFFPEVEIFLGRS 160
Db 124 PLIIVDEGTATTYCYINEHKQYMGALAPGIMISTEALFARAAKLPRIEIA RPDDIIGKN 183
QY 161 TRECVLGG---AYRESTEFIKSTLKLWRKVKFRKFKVITGGEGKYFSK----FGIYDPL 213
Db 184 TVSAMQAGILYGYVGQVEGIVSRMKAKSKI---PPKVIAATGGLAPLIASESDIIDVDPF 240
QY 214 LVHRGMRNLLY 224
Db 241 LTLTGLK-LLY 250
RESULT 12
AAU91158
ID AAU91158 standard; Protein; 262 AA.
XX
AC AAU91158;
XX
DT 05-JUN-2002 (first entry)
XX
DE Borrelia burgdorferi pantothenate kinase Coax.
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Borrelia burgdorferi.
```

```
XX WO200216601-A2.
PN
XX
XX 28-FEB-2002.
PD
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein -
XX
PS Claim 6; Page 76; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX
SQ Sequence 262 AA;
Query Match 12.1%; Score 141.5; DB 23; Length 262;
Best Local Similarity 28.3%; Pred. No. 4.2e-06;
Matches 62; Conservative 38; Mismatches 94; Indels 25; Gaps 9;
QY 4 LTVDVGNSSVDIALWEGKKVKDFLK-----LSHEE---FLKE--EFPKALKALGTSVKQS 52
Db 9 LIIDIGNTSIAFALFKDNQVNLFIKMTNLMRLRYDEVYSFFEENFDNFVNKVFISVVPI 68
QY 53 FSEKVRGKIPKIKFLKE---NFPIQVD----YKTPETLGTDRVALAYSAKKFFYK-K 101
Db 69 LNETFKNVI--FSFFKIKPLFIFGLDNLDTFNPYKSDKFLGSDVFANLVAIENYSFE 126
QY 102 NVVVISAGTALVIDLV-LEGFKGGFITLGLGKKLKILSDLAEGIPPEFFPEVEIFLGRS 160
Db 127 NVLVVDLGTACTIFAVSRQDGILGGIINSGLINFNLSLLDNAYLIKPKFPISTPNLLERT 186
QY 161 TRECVLGGAYRESTEFIKSTLKLWRKVKFRKFKVITGG 199
Db 187 TSGSVNSGLFYQKYKYLIEGVYRDIKQMYKKKFNLIITGG 225
RESULT 13
AAU91154
ID AAU91154 standard; Protein; 255 AA.
XX
AC AAU91154;
XX
DT 05-JUN-2002 (first entry)
XX
DE Geobacter sulfurreducens pantothenate kinase Coax.
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Geobacter sulfurreducens.
XX
PN WO200216601-A2.
XX
```

PD 28-FEB-2002.  
XX  
PF 24-AUG-2001; 2001WO-US26531.  
XX  
PR 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX  
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX  
PI Yocum RR, Patterson TA;  
XX WPI; 2002-269358/31.  
DR  
XX Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -  
XX  
PS Claim 10; Page 72-73; 128pp; English.  
XX The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.  
XX  
SQ Sequence 255 AA;  
Query Match 11.9%; Score 139.5; DB 23; Length 255;  
Best Local Similarity 22.1%; Pred. No. 6.5e-06;  
Matches 57; Conservative 47; Mismatches 113; Indels 41; Gaps 9;  
QY 4 LTVDVGNSSVDIALWEGKKVDFLKLSSHEE-----FLKEEFPKLKALGIS-----VK 50  
Db 3 LVIDVGNNTNIVLGIYDGERLVRDWRVSTDKARTTDEYGILINELFRLAGLDQIRAVII 62  
QY 51 QSFSEKVRGKIPKIF-----LK-----KENFPIQVDYKTPETLGTDRVALAYSAKKF 98  
Db 63 SSVVPPLTGVLERLSLGYFGMRPLVVGPGIKTGMPIQ--YDNPREVGDRIVNAVAGYEK 120  
QY 99 YGKNVVVISAGTALVIDLV-LEGKFKGGFITLGLGKKLKILSDLAEGIPEFFPEEVEIFL 157  
Db 121 YRTSLIIVDFGATTFFDYVNRKGEYCGGAIAPGLVISTEALFQRASKLPRVDIIRPSAII 180  
QY 158 GRSTRECVLGGAYRESTEFIKLWKRKVKFKKVVITG-----EGKYFSKFGIY 210  
Db 181 ARNTVNSMQAGIYYGYVGLVDEIVTRMKAESKDAPRVATGGLASLIAPESKTIEAVEEY 240  
QY 211 DPLLVRHGRMNLVLYHR 228  
Db 241 ---LTLEGLR---ILYER 252  
RESULT 14  
AAW98422  
ID AAW98422 standard; Protein; 223 AA.  
XX  
AC AAW98422;  
XX  
DT 31-MAR-1999 (first entry)  
XX  
DE H. pylori GHPO 344 protein.  
XX GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;  
KW peptic ulcer disease.  
XX Helicobacter pylori.  
OS  
XX WO9843478-A1.  
PN

XX 08-OCT-1998.  
PD  
XX  
PF 01-APR-1998; 98WO-US06371.  
XX  
PR 29-JUL-1997; 97US-0902615.  
PR 01-APR-1997; 97US-0833457.  
PR 24-JUN-1997; 97US-0881227.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
PA (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.  
XX  
PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;  
XX  
DR WPI; 1998-542293/46.  
DR N-PSDB; AAX14141.  
XX  
PT New isolated Helicobacter polynucleotides - used to develop products  
PT for the diagnosis, prevention and treatment of Helicobacter  
PT infections and gastrointestinal diseases  
XX  
PS Claim 8; Page 709-710; 2054pp; English.  
XX This sequence represents a Helicobacter pylori GHPO protein of the  
CC invention. The polypeptides can be used for preventing or treating  
CC Helicobacter infections, and gastroduodenal diseases associated with  
CC these infections, including acute, chronic, and atrophic gastritis, and  
CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be  
CC used for the production of antibodies. The products can also be used for  
CC detection and diagnosis.  
XX  
SQ Sequence 223 AA;  
Query Match 11.9%; Score 138.5; DB 19; Length 223;  
Best Local Similarity 26.2%; Pred. No. 6.8e-06;  
Matches 59; Conservative 35; Mismatches 88; Indels 43; Gaps 10;  
QY 7 DVGNSSVDIALWEGKKVDFLKLSSHEEFLKEEFPKLKALG-----ISVKQSFSEKVRG 59  
Db 17 DIGNTRIHA-----QNYQLFSSAKEDLKRGLGQKEIFYISVNEENEKALLN 63  
QY 60 KIPKIKFLKKENFPIQVDYKTPETLGTDRVALAYSAKKFYGNVVVISAGTALVIDLVLE 119  
Db 64 CYPNAKNI-AGFFHLETDY---VGLGIDRQMACLAV-----NNGVVVDAGSAITIDLIKE 114  
QY 120 GKFKGGFITLGLGKKLKILSDLAEGIPEFFP--EVEIFLGRSTRECVLGGAYRESTEFI 177  
Db 115 GKHLGGCILPGLAQYIHAYKKSAKILEQPFKALDSLEV-LPKSTRDVAENVG-----MV 166  
QY 178 KSTLKLWRKVKFRKFKVITGGEKGYFSKF---GIYDPLLVRHGM 219  
Db 167 LSVIACIQHLAKNQ-KIYLCGGDAKYLFAFLPHSVCKERLVDGM 210  
RESULT 15  
AAU91161  
ID AAU91161 standard; Protein; 223 AA.  
XX  
AC AAU91161;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Helicobacter pylori pantothenate kinase Coax.  
XX  
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX Helicobacter pylori.  
OS  
XX WO200216601-A2.  
XX  
PD 28-FEB-2002.  
XX

Search completed: June 24, 2003, 21:46:20  
Job time : 24.511 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:24:12 ; Search time 13.3935 Seconds  
(without alignments)  
1947.059 Million cell updates/sec

Title: US-09-813-453A-63  
Perfect score: 1264  
Sequence: 1 MIFIAVGNTRTLLAHTHDGV.....AGSLYQPHLLLRGFLWIRG 241

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1264	100.0	241	9	US-09-813-453A-63	Sequence 63, Appl
2	229.5	18.2	246	9	US-09-813-453A-9	Sequence 9, Appli
3	227	18.0	257	9	US-09-813-453A-53	Sequence 53, Appl
4	227	18.0	260	9	US-09-813-453A-51	Sequence 51, Appl
5	215	17.0	262	9	US-09-813-453A-8	Sequence 8, Appli
6	203	16.1	258	9	US-09-813-453A-2	Sequence 2, Appli
7	190.5	15.1	262	9	US-09-813-453A-45	Sequence 45, Appl
8	189	15.0	219	9	US-09-813-453A-57	Sequence 57, Appl
9	184	14.6	265	9	US-09-813-453A-4	Sequence 4, Appli
10	183.5	14.5	244	9	US-09-813-453A-41	Sequence 41, Appl
11	181	14.3	233	9	US-09-813-453A-17	Sequence 17, Appl
12	180.5	14.3	212	9	US-09-813-453A-59	Sequence 59, Appl
13	179.5	14.2	254	9	US-09-813-453A-47	Sequence 47, Appl
14	178.5	14.1	258	9	US-09-813-453A-49	Sequence 49, Appl
15	174	13.8	257	9	US-09-813-453A-13	Sequence 13, Appl
16	172.5	13.6	255	9	US-09-813-453A-7	Sequence 7, Appli
17	170.5	13.5	273	9	US-09-813-453A-10	Sequence 10, Appl
18	169	13.4	250	9	US-09-813-453A-3	Sequence 3, Appli
19	168.5	13.3	258	9	US-09-813-453A-6	Sequence 6, Appli

20	164.5	13.0	229	9	US-09-813-453A-12	Sequence 12, Appl
21	159	12.6	272	9	US-09-712-363-276	Sequence 276, App
22	159	12.6	272	9	US-09-813-453A-5	Sequence 5, Appli
23	154.5	12.2	267	9	US-09-813-453A-15	Sequence 15, Appl
24	145	11.5	256	9	US-09-813-453A-55	Sequence 55, Appl
25	144.5	11.4	223	9	US-09-895-913A-74	Sequence 74, Appl
26	144.5	11.4	223	9	US-09-813-453A-14	Sequence 14, Appl
27	144.5	11.4	223	9	US-09-813-453A-67	Sequence 67, Appl
28	144.5	11.4	248	9	US-09-813-453A-20	Sequence 20, Appl
29	143.5	11.4	242	9	US-09-813-453A-65	Sequence 65, Appl
30	136	10.8	262	9	US-09-813-453A-11	Sequence 11, Appl
31	124	9.8	209	9	US-09-813-453A-21	Sequence 21, Appl
32	123.5	9.8	460	9	US-09-813-453A-39	Sequence 39, Appl
33	115.5	9.1	592	9	US-09-813-453A-43	Sequence 43, Appl
34	112.5	8.9	592	9	US-09-813-453A-22	Sequence 22, Appl
35	107.5	8.5	249	9	US-09-813-453A-70	Sequence 70, Appl
36	104.5	8.3	249	9	US-09-813-453A-61	Sequence 61, Appl
37	94.5	7.5	1827	9	US-09-712-363-261	Sequence 261, App
38	94.5	7.5	5215	9	US-09-860-846-2	Sequence 2, Appli
39	94.5	7.5	5215	9	US-09-988-384B-2	Sequence 2, Appli
40	94.5	7.5	5215	9	US-09-836-821-2	Sequence 2, Appli
41	94.5	7.5	5215	10	US-09-861-289-2	Sequence 2, Appli
42	89	7.0	507	9	US-10-270-333-45	Sequence 45, Appl
43	86	6.8	732	9	US-10-138-927-78	Sequence 78, Appl
44	86	6.8	1077	9	US-10-121-911-1	Sequence 1, Appli
45	85	6.7	760	9	US-09-712-363-292	Sequence 292, App

ALIGNMENTS

RESULT 1  
US-09-813-453A-63  
; Sequence 63, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 63  
; LENGTH: 241  
; TYPE: PFT  
; ORGANISM: Thiobacillus ferrooxidans  
US-09-813-453A-63

Query Match	100.0%	Score 1264	DB 9	Length 241
Best Local Similarity	100.0%	Pred. No. 6.1e-113		
Matches 241	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Qy	1	MIFIAVGNTRTLLAHTHDGVHSDSVATSLPPTTEILQQPGLTWLSAPNREPVALGGVVP	60	
Db	1	MIFIAVGNTRTLLAHTHDGVHSDSVATSLPPTTEILQQPGLTWLSAPNREPVALGGVVP	60	
Qy	61	AALAAWREALATAEVREPDGFFRRAVPHDYHPPESLGFDRRCCLLAAMDYPGQDSIVI	120	
Db	61	AALAAWREALATAEVREPDGFFRRAVPHDYHPPESLGFDRRCCLLAAMDYPGQDSIVI	120	
Qy	121	DMGTAITIDLLAGGHFRGGRILPGIAMSRLRGLHEGTALLPEVVLNAPAEMLGNDTSNAIQ	180	
Db	121	DMGTAITIDLLAGGHFRGGRILPGIAMSRLRGLHEGTALLPEVVLNAPAEMLGNDTSNAIQ	180	
Qy	181	AGVIHLFADALRGAITDFRQYSPQARILITGGDAERWQPGIAGSLYQPHLLLRGFLWIR	240	
Db	181	AGVIHLFADALRGAITDFRQYSPQARILITGGDAERWQPGIAGSLYQPHLLLRGFLWIR	240	

Db 181 AGVIHLFADALRGAITDFRQYSPQARILITGGDAERWQPGIAGSLYQPHLLLRGFYLMWR 240

QY 241 G 241

Db 241 G 241

RESULT 2

US-09-813-453A-9

; Sequence 9, Application US/09813453A

; Patent No. US20020168681A1

; GENERAL INFORMATION:

; APPLICANT: Yocum, R. Rogers

; APPLICANT: Patterson, Thomas A.

; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF

; TITLE OF INVENTION: ANTIBIOTICS

; FILE REFERENCE: OGZ-001

; CURRENT APPLICATION NUMBER: US/09/813,453A

; CURRENT FILING DATE: 2001-03-20

; PRIOR APPLICATION NUMBER: US 60/227,860

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: 09/667,569

; PRIOR FILING DATE: 2000-09-21

; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 9

; LENGTH: 246

; TYPE: PRT

; ORGANISM: Thermotoga maritima

US-09-813-453A-9

Query Match 18.2%; Score 229.5; DB 9; Length 246;

Best Local Similarity 28.2%; Pred. No. 4e-14;

Matches 71; Conservative 40; Mismatches 110; Indels 31; Gaps 5;

QY 2 IFIAGVNTRTLAAHTDGVHFDVSVVATSLPPTTEILQQPGL-TWLSAPNREPVALG--GV 58

Db 3 LLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDELEFSLHPLLLGDAMREIKGIGVASV 62

QY 59 VPAALAA-----WREALATAEVREPDPGFFRRVPHDYHPPESLGFDRRC 104

Db 63 VPTQNTVIERFSQYFHHISPIWKAK-----NGCVKNVKNPSEVGADRVAN 109

QY 105 LLAAMDYPGQDSIVIDMGTAITIDLLAGHFRGGRILPGIAMSRLRGLHEGTALLPEVVL 164

Db 110 VVAFVKEY-GKNGIIDMGTAITVDLVVNGSVYEGGAILPGFFMMVHSLFRGTAKPLVEV 168

QY 165 NPAEMLGNDTSNAIQAGVIHLFADALRGAITDFRQYSPQARILITGGDAERWQPGIAGS 224

Db 169 KPADFVVGKDTFENIRLGVVNGSVYALEGIIGRIKEVYGDLPVLTGGOSKIVKDMIKHE 228

Y 225 LYQPHLLLRGFY 236

Db 229 IFDEDLTIKGVY 240

RESULT 3

US-09-813-453A-53

; Sequence 53, Application US/09813453A

; Patent No. US20020168681A1

; GENERAL INFORMATION:

; APPLICANT: Yocum, R. Rogers

; APPLICANT: Patterson, Thomas A.

; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF

; TITLE OF INVENTION: ANTIBIOTICS

; FILE REFERENCE: OGZ-001

; CURRENT APPLICATION NUMBER: US/09/813,453A

; CURRENT FILING DATE: 2001-03-20

; PRIOR APPLICATION NUMBER: US 60/227,860

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: 09/667,569

; PRIOR FILING DATE: 2000-09-21

; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 53

; LENGTH: 257

; TYPE: PRT

; ORGANISM: Chlorobium tepidum

US-09-813-453A-53

Query Match 18.0%; Score 227; DB 9; Length 257;

Best Local Similarity 32.1%; Pred. No. 7.3e-14;

Matches 84; Conservative 43; Mismatches 101; Indels 34; Gaps 12;

QY 2 IFIAGVNTRTLAAHTDGVHFDVSVVATSLPPTTEILQQPGLTWLSAPNREP--VALGGV 58

Db 3 LVVDIGNTSTTLA-IFTGDEEPSVESVPSALFADSTMRVFCNMARKHGEPOAIAICSV 61

QY 59 VPAALAA---AWREALATAEV-----REPDGFFRRVPHDYHPPESLGFDRRCCLLAA 108

Db 62 VPSATAVGSALLESLFSVPVLTICCKLRFP-----FRL----DYATPHTFGADRLAICAWS 113

QY 109 AMDYPGQDSIVIDMGTAITIDLL-AGHFRGGRILPGIAMSRLRGLHEGTALLPEVVLNAP 167

Db 114 RHLFSEKPVIAVDIGTAITFDVLDTVGNRYGGLIMPIDMAGALHSRTAQLPOVRIDRP 173

QY 168 AEMLGNDTSNAIQAGV-----IHLFADALRG-AITDFRQYSPQARILITGGDAERWQ 219

Db 174 ESSLGRSTTECIKSGVFWGVVQIGGLVDAIRGLDVRDFGEST--VEVITGGNSRIIVP 231

QY 220 GIAG-SLYQPHLLLRGFYLMWR 240

Db 232 EIGPVSVIDEAVLRGSDLLLR 253

RESULT 4

US-09-813-453A-51

; Sequence 51, Application US/09813453A

; Patent No. US20020168681A1

; GENERAL INFORMATION:

; APPLICANT: Yocum, R. Rogers

; APPLICANT: Patterson, Thomas A.

; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF

; TITLE OF INVENTION: ANTIBIOTICS

; FILE REFERENCE: OGZ-001

; CURRENT APPLICATION NUMBER: US/09/813,453A

; CURRENT FILING DATE: 2001-03-20

; PRIOR APPLICATION NUMBER: US 60/227,860

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: 09/667,569

; PRIOR FILING DATE: 2000-09-21

; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 51

; LENGTH: 260

; TYPE: PRT

; ORGANISM: Caulobacter crescentus

US-09-813-453A-51

Query Match 18.0%; Score 227; DB 9; Length 260;

Best Local Similarity 29.0%; Pred. No. 7.4e-14;

Matches 75; Conservative 43; Mismatches 103; Indels 38; Gaps 12;

QY 1 MIFIAGVNTRTLAAHTDGV-----VHFDSVSVATSLPP-----TEILQQPGLTWLSAPNR 50

Db 2 LLAIEQGNNTMTFA-IHDGASWVAQWRSATESRTRTADEVVWLSQLLSMOGLGFRAI--- 57

QY 51 EPVALGGVVPAAALAAWR-----EALATAEVREPDPGFFRRVPHDYHPPESLGFDR 101

Db 58 DAVIISVVVPOQSIFNLRLNLSRRYFNVEPLVIGENA-----KLGDVRIEKPSEAGADR 110

QY 102 RCCLLAAAMDYPGQDSIVIDMGTAITIDLLAG-GHFRGGRILPGIAMSRLRGLHEGTALLP 160

Db 111 LVNAIGAAMVYPG-PLVVIDSGTATTFDIVAADGAFEGGIAPGINLSMQALHEAAAKLP 169

QY 161 EEWLNAPA--EMLGNDTSNAIQAGVIHLFADALRGAITDFR-QYSPQARILITGGDAERW 217









```

; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 49
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Bacillus stearothermophil
US-09-813-453A-49

```

Query Match	14.1%;	Score	178.5;	DB	9;	Length	258;
Best Local Similarity	28.0%;	Pred.	No. 3.2e-09;				
Matches	73;	Conservative	42;	Mismatches	109;	Indels	37; Gaps
							14;
QY	1	MIFIA-VGNTRFLLAHTHDGVHFDVSV----	VATSLPPT-----ILQQPGLTWLS	46			
Dd	1	MIFVLVDGNTNVLG-VYDG---	DELKHHWRIFTSRKTEDGYGMMIKALLNHVGJQF--	54			
QY	47	APNREPVALGGVYPAAAL-AAWREALATAEVRE--	PDGFFERRAVPHDYHPPESELGFDRRC	103			
b	55	-SDIRGIISVVPPIMFALERMCIFYFHIKPLIVGP-	IKTGIDIKYDNPREVGA DRV	112			
QY	104	CLLAAAMDYPGDSIVIDMGTAITIDLL-AGCHFRGRILPGIAMSRLRGHEGTALLPEV	162				
Dd	113	NAVAGIHLY-GSPLIIIVDFGTATTTCYINEHKQYMGGAIPGIMISTEALFARA AKLPRI	171				
QY	163	VLNAPAEMLGNDTSAIQAGVIHLFADALRGAITDFROYSP-QARILITGGDAERWPQGI	221				
Dd	172	EIARPD DIIGNKTVSAMOAGILYG YVGQVEGIVSRMKAKSKIPPKV IATGGLA----	PLI 227				
QY	222	AG-----SLYQPHLLL RGFYL	237				
Dd	228	ASESDIIDVVDVPFLTLTGLKL	248				

RESULT 15  
US-09-813-453A-13  
; Sequence 13, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 13  
; LENGTH: 257  
; TYPE: PRT  
; ORGANISM: Synecchocystis sp.  
US-09-813-453A-13

	Query Match	13.8%;	Score 174;	DB 9;	Length 257;
	Best Local Similarity	27.9%;	Pred. No. 8.4e-09;		
	Matches 60;	Conservative 34;	Mismatches 73;	Indels 48;	Gaps 8;
yy	40	PGLTWLSAPN-----	REPVALGGVVPAALAAWREALATAEVR	PDGFFRR	AVP 88
		::	:  ::		:
Db	40	PLQTVWTDYNPKSAQLPVLLGKVPLMLASV	VEQTEVWR-----	VIQPKILTLK	NLPL 92
yy	89	HDYHPPESLGFDRRCCLLAAAMDYPGQDS	IVIDMGTAITIDLLAGG	HFR-----	GGRILP 143
		: :	:     :		:
Db	93	VNLYP--SFGIDRALAGLGTGLTY-GFPCL	VVDGGTALTIT----	GFDQDKKL	VGGAILP 145
yy	144	GIAMSLRGLHGGTALLPEVVLNAPAEMLGN	----DTSNAIOAGVIHLFADALRG	AITD	FERO 200

[illegible]



Db 3 LLVDVGNTHSVFSITEDGKTFRRWRRLSTGVFQTEDELFSHLPLGLGDAMREIKGIGVASV 62  
QY 59 VPAALAA-----WREALATAEVRPDPGFFERRAVPHDYHPPESLGFDRCC 104  
Db 63 VPTQNTVIERFSQYFHIPIWKAK-----NGCVKWNVNKPNSEVGADRVAN 109  
QY 105 LLAAMDYPGQDSIVDMGTAITIDLLAGHFRGGRILPGIAMSRLGLHEGTALLPEVVL 164  
Db 110 VVAFVKEY-GKNGIIDMGTAITTDLVVNGSYEGGAILPGFFMMVHSLFRGTAKLPLEV 168  
QY 165 NAPAEMLGNDTSNAIQAGVIHLFADALRGAITDFRQYSPQARILITGGDAERWQPGIAGS 224  
Db 169 KPADFVVGKDTENIRLGVVNGSVYALEGIIGRIKEVYGDLPVVLVTGGQSKIVKDMIKHE 228  
QY 225 LYQPHLLLRGFY 236  
Db 229 IFDEDLTIKGVY 240

RESULT 2  
Q9A6Z1

Q9A6Z1 PRELIMINARY; PRT; 261 AA.  
Q9A6Z1;  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Transcriptional activator, putative, Baf family.  
GN CC1935.  
OS Caulobacter crescentus.  
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
OC Caulobacter.  
OX NCBI\_TaxID=155892;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 19089 / CB15;  
RX MEDLINE=21173698; PubMed=11259647;  
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,  
Kolony J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,  
Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,  
Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
RT "Complete genome sequence of Caulobacter crescentus."  
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
DR EMBL; AE005867; AAK23910.1; -  
DR TIGR; CC1935; -  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMS; TIGR00671; baf; 1.  
QY Complete proteome.  
QY SEQUENCE 261 AA; 27965 MW; C19E60D7B0714EF5 CRC64;

Query Match 18.0%; Score 228; DB 16; Length 261;  
Best Local Similarity 29.1%; Pred. No. 1.le-10;  
Matches 76; Conservative 44; Mismatches 101; Indels 40; Gaps 13;

QY 1 MIFIAV--GNRTLLAHTHDG---VHEDSVSVATSLPP-----TEILQOPGLTWLSAP 48  
Db 1 MLLAIEQGNNTMTFA-IHDGASVWAQWRSATESRTRADEYVWLSQLLSMQGLGFRAI- 58  
QY 49 NREPVALGGVVPAALAAWR-----EALATAEVRPDPGFFRRAVPHDYHPPESLGF 99  
Db 59 --DAVISSVVPQSFNLRNLSRRYFNVEPLVIGENA-----KLGIDVRIEKPSEAGA 109  
QY 100 DRRCCLLAAAMDYPGQDSIVDMGTAITIDLLAG-GHFRGGRILPGIAMSRLGLHEGTAL 158  
Db 110 DRLVNAIGAAMVPG-PLVIDSGTATFDIVAADGAFEGGIIAPGINSMQALHEAAK 168  
QY 159 LPEVVLNAPA--EMLGNDTSNAIQAGVIHLFADALRGAITDFR-QYSPQARILITGGDAE 215  
Db 169 LPRIAIOQPAGNRIVGTDVTSAMQSGVFWGYSISLIEGLVARIKAEERGEPMTVIATGGVAS 228

QY 216 RWQPGIAGSL--YQPHLLLRG 234  
Db 229 LFE-GATDSIDHFDSDLTIRG 248  
RESULT 3  
Q9RX54  
ID Q9RX54 PRELIMINARY; PRT; 262 AA.  
AC Q9RX54;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Hypothetical protein DR0461.  
GN DR0461.  
OS Deinococcus radiodurans.  
OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;  
OC Deinococcaceae; Deinococcus.  
OX NCBI\_TaxID=1299;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=R1;  
RX MEDLINE=20036896; PubMed=10567266;  
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,  
Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
Fraser C.M.;  
RT "Genome sequence of the radioresistant bacterium Deinococcus  
radiodurans R1."  
RL Science 286:1571-1577(1999).  
DR EMBL; AE001905; AAF10040.1; -  
DR TIGR; DR0461; -  
DR InterPro; IPR004619; Baf.  
DR InterPro; IPR001220; Lectin\_legB.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMS; TIGR00671; baf; 1.  
DR PROSITE; PS00307; LECTIN\_LEGUME\_BETA; UNKNOWN\_1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 262 AA; 27839 MW; 965EAAD2F78785A0 CRC64;

Query Match 17.0%; Score 215; DB 16; Length 262;  
Best Local Similarity 29.7%; Pred. No. 1.2e-09;  
Matches 78; Conservative 35; Mismatches 112; Indels 38; Gaps 11;

QY 1 MIFIAVGNTRTLL-----AHTDGVHFDVSVSVATS---LPPTEILQOPGL-TWLSAPN 49  
Db 6 LLAVDIGNTTTLVLGLADASGALTH-----TWIRTNREMLPDDLALQLHGLFTLAGAPI 59  
QY 50 REPVALGGVVPA-----ALAAWR-----EALATAEVRPDPGFFRRAVPHDYHPPESLGF 100  
Db 60 PRAAVLSSVAPPVGENYALAKRHFMDAFVSAENLPD-----VTVELDTPGSVGAD 112  
QY 101 RRCCLLAAAMDYPGQD-SIVDMGTAITIDLLA-GGHRGGRILPGIAMSRLGLHEGTAL 158  
Db 113 RLCNLFGAEKYLGGLDYAVVDFGTSTNFDVVGRRRFLGGLATGAQVSADALFARAAK 172  
QY 159 LPEVVLNAPAEMLGNDTSNAIQAGVIHLFADALRGAITDFRQYSPQARILITGGDAERWQ 218  
Db 173 LPRITLQAPETAIGKNTVHALQSLVFGVYAEWVDGLLRIRRAELPGEAVAVATGGFSRTV 232  
QY 219 PGIAGSL--YQPHLLLRGFY-LW 238  
Db 233 QGICQEIYDYDETTLRLGLVELW 255

RESULT 4  
Q8RFE4

ID Q8RFE4 PRELIMINARY; PRT; 256 AA.  
AC Q8RFE4;  
DT 01-JUN-2002 (TREMBlrel. 21, Created)  
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)



DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Bvg accessory factor.  
GN FN0761.  
OS Fusobacterium nucleatum (subsp. nucleatum).  
OC Bacteria; Fusobacteria; Fusobacterium.  
OX NCBI\_TaxID=76856;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 25586;  
RX MEDLINE=21886394; PubMed=11889109;  
RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,  
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,  
RA Vasieva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,  
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,  
RA Fonstein M., Kyrpides N., Overbeek R.;  
RT "Genome sequence and analysis of the oral bacterium Fusobacterium  
RT nucleatum strain ATCC 25586.";  
RL J. Bacteriol. 184:2005-2018(2002).  
RW EMBL; AE010586; AAL94957.1; -.  
SQ SEQUENCE 256 AA; 28601 MW; CDF681127F08914B CRC64;  
Query Match 16.1%; Score 203; DB 16; Length 256;  
Best Local Similarity 25.5%; Pred. No. 1.1e-08;  
Matches 67; Conservative 44; Mismatches 110; Indels 42; Gaps 8;  
QY 1 MIFIAVGNTRELLA-HTHDGVHFDVSVATSLPPTPE-----ILQQPGLTWLSAPNREP 53  
DB 2 IIGIDIGNTHVTGIYDNGELISTFRIATNDKMTDEYFSYFNITKYNEISIKKVD 61  
QY 54 ALGGVVPALAAWR-----EALATAVEPREDPGFFRRAVPHDYHPHESLGFD 104  
DB 62 LISSVVPNIITFFQFFARKYFKVEA-TIVDLKELPFTFAKGINY-----TGFGADRI 115  
QY 105 LLAAMDYPGQDSIVIDMGTAITIDLLAGHFRGGRILPGIAMSRLRGLHGTALLPE 164  
DB 116 ITEAMQKYPDKNLVIFDFGTATYDVLKKGVIYGGILPGIDMSINALYGN 175  
QY 165 NAPAEMLGNDTSNAIQAGVIHLFADALRGAITDF-RQYSPQARILITGGDAERWQPGI 223  
DB 176 TTPSSVLGTDPMKQIQAAIFFGYAGQIKHIKKINEELNEEIFVLATGG-----LG 226  
QY 224 SL-----YQPHLLLRGFY 236  
DB 227 KILSAEIDEIDEYDANLSLKG 249  
RESULT 5  
QY Q8YQD7 PRELIMINARY; PRT; 276 AA.  
AC Q8YQD7;  
DT 01-MAR-2002 (TReMBLrel. 20, Created)  
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Hypothetical protein Alr3896.  
GN ALR3896.  
OS Anabaena sp. (strain PCC 7120).  
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
OX NCBI\_TaxID=103690;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21595285; PubMed=11759840;  
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,  
RA Yasuda M., Tabata S.;  
RT "Complete genomic sequence of the filamentous nitrogen-fixing  
RT cyanobacterium Anabaena sp. strain PCC 7120.";  
RL DNA Res. 8:205-213(2001).  
DR EMBL; AP003594; BAB75595.1; -.  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.

KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 276 AA; 30272 MW; 2AD04CA693E56F25 CRC64;  
Query Match 15.0%; Score 190; DB 16; Length 276;  
Best Local Similarity 25.6%; Pred. No. 1.3e-07;  
Matches 72; Conservative 35; Mismatches 108; Indels 66; Gaps 11;  
QY 4 IAVGNTR-----TLAHTHDGVHFDVSVATSLPPTTEILQQPG----- 41  
DB 17 LEIGNSRLHWALFMGESLEFTWD-----TEYLPESVIOQLNGENGETKLEVGSEKE 66  
QY 42 --LTWL---SAPNREPVALGGVVPAAALAAWREALATAEVRPDGFFERRAVPHDYHPPE- 95  
DB 67 IFFTFPLPPAPCPLPLFIASVVPQTVLWENYLVN-----RVITLDQIPLNN 114  
QY 96 ---SLGFDRRCCLLAAAMDYPGQDSIVIDMGTAITIDLLAGG-HFRGGRILPGIAMSRLG 151  
DB 115 IYPTLGIDRALALWGAGMSW-GFPVLVIDAGTALTFTAADGKNLVGGAILPGVGLQFAS 173  
QY 152 LHEGTALLPEVVLNAPAE---LGNDTSNAIQAGVIHLFADALRGAITDFRQYSPQARIL 208  
DB 174 LGQQTGQLPQVEMEAIKSLPPRFALNTTEAIQSGVIYTTIAGMRDFTTEWLSLFPDGKVA 233  
QY 209 ITGGD-----AERWQPGIAGSL-YOPHLLLRGFYLWIRG 241  
DB 234 IKGGDRILLNLYLQALYPLDLAARLIVEPNLIFWGMQTVAG 274  
RESULT 6  
QY Q8YAC5 PRELIMINARY; PRT; 259 AA.  
AC Q8YAC5;  
DT 01-MAR-2002 (TReMBLrel. 20, Created)  
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Hypothetical protein lmo0221.  
GN LMO0221.  
OS Listeria monocytogenes.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Listeriaceae; Listeria.  
OX NCBI\_TaxID=1639;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=EGD-E / SEROVAR 1/2A;  
RX MEDLINE=21537279; PubMed=11679669;  
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
RA Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,  
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,  
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,  
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,  
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,  
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;  
RT "Comparative genomics of Listeria species.";  
RL Science 294:849-852(2001).  
DR EMBL; AL591974; CAD00748.1; -.  
DR MEROPS; M41.009; -.  
DR ListiList; LMO00221; -.  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGREMS; TIGR00671; baf; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 259 AA; 28187 MW; F106B049D80025B7 CRC64;  
Query Match 14.6%; Score 185; DB 16; Length 259;  
Best Local Similarity 27.0%; Pred. No. 3.1e-07;  
Matches 72; Conservative 43; Mismatches 108; Indels 44; Gaps 12;  
QY 1 MIFIAVGNT-----RTLLAH---THDGVHFDVSVATSLPPTTEILQQPGLTWLSAP 48  
DB 2 IIGIDIGNTHVTGIYDNGELISTFRIATNDKMTDEYFSYFNITKYNEISIKKVD 61

Db 2 ILVIDVGNCTVGVYKQKLLKHWMTTDR-HRTSDELGMTV-----LNFFSYA 50  
QY 49 NREP-----VALGGVVPVPAALAAWREALATAEVR---EP---DPGFFRAVPHDYHPPES 96  
Db 51 NLTPSDIOGIISSVVPIM---HAMETMCVRYFNIRPLIVPG-IKTGLNLKVDNPRE 105  
QY 97 LGFDRRCCLLAAAMDYPGQDSIVIDMGTAITIDLL-AGGHFRGGRILPGIAMSLRGLHEG 155  
Db 106 IGSDRVNAVAASEEY-GTPVIVVDFGTATTCYIDESGVYQGGAIAPGIMISTEALYNR 164  
QY 156 TALLPEVVLNAPAEMLGNDTSNAIQAGVIHLFADALRGAITDFRQYSPQARILITGGDAE 215  
Db 165 AAKLPRVDIAESSQIIIGKSTVSSMQAGIFYGVGQCEGIIAEMKKQSNASPVVATGGLA 224  
QY 216 RW--QPGIAGSLYQPHLLLRGFYLR 240  
Db 225 RMITEKSSAVDILDPFELTKGLELLYR 251

RESULT 7

Q9X8N6 PRELIMINARY; PRT; 265 AA.  
Q9X8N6;  
DT 01-NOV-1999 (TremBLrel. 12, Created)  
DT 01-NOV-1999 (TremBLrel. 12, Last sequence update)  
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)  
DE Hypothetical protein SCO3380.  
GN SCO3380 OR SCE94.31C.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Oliver K., Harris D.;  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RX MEDLINE=97000351; PubMed=8843436;  
RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,  
RA Kinashi H., Hopwood D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for  
the 8 Mb Streptomyces coelicolor A3(2) chromosome."  
Mol. Microbiol. 21:77-96(1996).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2) / M145;  
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,  
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
RA Hopwood D.A.;  
RT "Complete genome sequence of the model actinomycete Streptomyces  
coelicolor A3(2).";  
RL Nature 417:141-147(2002).  
DR EMBL; AL049628; CAB40880.1; -.  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMS; TIGR00671; baf; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 265 AA; 28203 MW; E34FA2ABBA946969 CRC64;

Query Match 14.6%; Score 184; DB 16; Length 265;  
Best Local Similarity 28.8%; Pred. No. 3.8e-07;  
Matches 76; Conservative 39; Mismatches 113; Indels 36; Gaps 11;  
QY 1 MIFIAVGNTRTLAHTHDG-----VHFDSSVSVATSLPPTTEILQ-----QPGLTWLSA 47  
Db 2 LLTIDVGNTHTVLG-LFDGEDIVEHWRISTDSRRTADEL--AVLQGLMGMLPLLGDELG 58  
QY 48 PNREPVALGGVVPVPAALAAWREALATAEVR---DPGFERRAVPHDYHPPESLGFDRRC 103  
Db 59 DGIDGIAICATVPSVLHRELVTRRYGDPVAVLVEPG-VKTGVPIILTDHPKEVGADRII 117  
QY 104 CLLAAAMDYPGQDSIVIDMGTAITIDLL-AGGHFRGGRILPGIAMSLRGLHEGTALLPEV 162  
Db 118 NAVAARELY-GGPAIVVDFGTATTFDAVSARGEYIGGVIAPIGIEISVEALGVKGAQLRKI 176  
QY 163 VLNAPAEMLGNDTSNAIQAGVIHLFADALRGAIT---DFRQYSPQARILITGGDAERWQ 218  
Db 177 EVARPRSVIGKNTVEAMQSGIVYGFAGQVDGVVNRMAELADDDPDDVTVIATGGLA---- 232  
QY 219 PGIAGSL-----YQPHLLLRGFYL 237  
Db 233 PMVLGESSVIDEHEPWLTMGLRL 256

RESULT 8

Q92F54 PRELIMINARY; PRT; 259 AA.  
AC Q92F54;  
DT 01-DEC-2001 (TremBLrel. 19, Created)  
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)  
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)  
DE Hypothetical protein lin0253.  
GN Lin0253.  
OS Listeria innocua.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Listeriaceae; Listeria.  
OX NCBI\_TaxID=1642;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CLIP 11262 / SEROVAR 6A;  
RX PubMed=11679669;  
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,  
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,  
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
RA Jones L.-M., Kaverst U., Krefit J., Kuhn M., Kunst F., Kurapkat G.,  
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,  
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;  
RT "Comparative genomics of Listeria species."  
RL Science 294:849-852(2001).  
DR EMBL; AL596164; CAC95486.1; -.  
DR ListList; LIN00253; -.  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMS; TIGR00671; baf; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 259 AA; 28227 MW; 554B03A0C0EFA64F CRC64;

Query Match 14.3%; Score 181; DB 16; Length 259;  
Best Local Similarity 27.0%; Pred. No. 6.5e-07;  
Matches 72; Conservative 42; Mismatches 109; Indels 44; Gaps 12;  
QY 1 MIFIAVGNTRTLAHTHDG-----VHFDSSVSVATSLPPTTEILQ-----QPGLTWLSA 48  
Db 2 LLTIDVGNTHTVLG-LFDGEDIVEHWRISTDSRRTADEL--AVLQGLMGMLPLLGDELG 58  
QY 49 NREP-----VALGGVVPVPAALAAWREALATAEVR---EP---DPGFFRAVPHDYHPPES 96

Db 51 NLTPSDIQIIGIISSVWPPIM---HAMETMCVRYFNIRPLIVPG-IKTGLNLKVDNPRE 105  
QY 97 LGFDRRCCLLAAAMDYPGQDSIVIDMGTAITIDLL-AGGHRGGRILPGIAMSRLRGLHEG 155  
Db 106 IGSDRIVNAVAASEEY-GTPVIVVDFGTATTCYIDEAGVYQGGAIAPGIMISTEALYNR 164  
QY 156 TALLPEVVLNAPAEMLGNDTSNAIQAGVIHLFADALRGAITDFRQYSPQARILITGGDAE 215  
Db 165 AAKLPRVDIAESSQIIGKSTVASMQAGIFYGFIGQCEGIIAEMKKQSNTPSPVVATGGLA 224  
QY 216 RW--QPGIAGSLYQPHLLLRGFYLWIR 240  
Db 225 RMIKSSAVDILDPPFLTKGLELLYR 251

RESULT 9  
O32514 PRELIMINARY; PRT; 212 AA.  
O32514;  
01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Similar to Bacillus subtilis.  
OS Desulfovibrio vulgaris (strain Miyazaki).  
OC Bacteria; Proteobacteria; delta subdivision; Desulfovibrionaceae;  
OC Desulfovibrio.  
OX NCBI\_TaxID=883;  
[1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=MIYAZAKI;  
RA Kitamura M., Konishi T., Kawanishi K., Ohashi K., Kishida Y.,  
RA Kohno K., Akutsu H., Kumagai I., Nakaya T.;  
RT "Sequence analyses of two ferredoxin genes and their flanking regions  
RT from Desulfovibrio vulgaris (Miyazaki F).";  
RL J. Biochem. Mol. Biol. Biophys. 2:147-154(1998).  
DR EMBL; AB005550; BAA21476.1; -.  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
SQ SEQUENCE 212 AA; 22336 MW; 23E1789973A344D7 CRC64;

Query Match 14.3%; Score 180.5; DB 2; Length 212;  
Best Local Similarity 29.4%; Pred. No. 5.6e-07;  
Matches 63; Conservative 31; Mismatches 83; Indels 37; Gaps 9;  
QY 6 VGNTRTLAHTHDGVHFDVSVATSLPPTTEILQQPGLTWLSAPNR--EPVALGGVVPAAAL 63  
Db 10 IGNTNVKI-----GIAVETAVLTSVLPDTPD---PGQTTDSIGLRLEVLRHAGLGPADV 60  
QY 64 AAWREALATAEVRPDPGFFRR-----AVPHD--YHPPESLGFDRRC 104  
Db 61 GA---CVASSVWPGVNP-LIRRACERYLYRKLFPAGDIAIPLDNYERPAEVDRLVA 116  
QY 105 LLAAMDYPGQDSIV-IDMGTAITIDLLAGGHRGGRILPGIAMSRLRGLHEGTALLPEVV 163  
Db 117 AYAARRLYPGRSLVSVDFGTATTFDCVEGGAYLGLICPGVLSSAGALSSRTAKLPRIS 176  
QY 164 LNAPAE--MLGNDTSNAIQAGVIHLFADALRGAI 195  
Db 177 LEVEDSPVIGRSTTTSLNHGFIQFAAMTEGVL 210

RESULT 10  
Q9F985 PRELIMINARY; PRT; 258 AA.  
AC Q9F985;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Putative 32 kDa replication protein.  
OS Bacillus stearothermophilus.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Geobacillus.  
OX NCBI\_TaxID=1422;

[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=V;  
RC Vasquez C., Pichuantes S., Saavedra C.;  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBSJ databases.  
DR EMBL; AF198621; AAG28531.1; -.  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMS; TIGR00671; baf; 1.  
SQ SEQUENCE 258 AA; 28101 MW; 507B55D695095855 CRC64;  
Query Match 14.3%; Score 180.5; DB 2; Length 258;  
Best Local Similarity 27.9%; Pred. No. 7.1e-07;  
Matches 72; Conservative 43; Mismatches 106; Indels 37; Gaps 14;  
QY 1 MIFIA-VGNTRTLAHTHDGVHFDVSVS---VATSLPPTTE-----ILQQPGLTWLS 46  
Db 1 MIFVLDVGNTNTVLG-VYDG---DELKHHWRIETSRGKTEDEYGMTIKALLNHVGLQF-- 54  
QY 47 APNREPVALGGVVPAAAL-AAWREALATAEVR--PDPGFFERRAVPHDYHPPESLGFDRRC 103  
Db 55 -SDIDGIISSVVPPIFALERMCKLYFHKPIIVGPG-IKTGLNIKYDNPREGADRIV 112  
QY 104 CLLAAAMDYPGQDSIVIDMGTAITIDLL-AGGHRGGRILPGIAMSRLRGLHEGTALLPEV 162  
Db 113 NAVAGIHLV-GSPLIIVDFGTATTTCYINEHKQYMGGAIPGIMISTEALFARAACLPR 171  
QY 163 VLNAPAEMLGNDTSNAIQAGVIHLFADALRGAITDFRQYSP-QARILITGGDAERWQPGI 221  
Db 172 EIARPDIIIGKNTVVSAMQAGILYGVGVQVEGIVSRMKAKSPVPPKVIATGGLASL----I 227  
QY 222 AG-----SLYQPHLLLRG 234  
Db 228 ASESNVIDIVDPFLTLTG 245

RESULT 11  
Q8Y2M4 PRELIMINARY; PRT; 295 AA.  
ID Q8Y2M4  
AC Q8Y2M4;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Putative transcription regulation accessory factor transcription  
DE regulator protein.  
GN RSC0311 OR RS03278.  
OS Ralstonia solanacearum (Pseudomonas solanacearum).  
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
OC Ralstonia.  
OX NCBI\_TaxID=305;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GM11000;  
RX MEDLINE=21681879; PubMed=11823852;  
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,  
RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,  
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,  
RA Sigulier P., Thebault P., Whalen M., Wincker P., Levy M.,  
RA Weissenbach J., Boucher C.A.;  
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";  
RL Nature 415:497-502(2002).  
DR EMBL; AL646058; CAD13839.1; -.  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
KW Complete proteome.  
SQ SEQUENCE 295 AA; 30990 MW; 133074A7764BFCBB CRC64;

Query Match 14.2%; Score 180; DB 16; Length 295;  
Best Local Similarity 26.2%; Pred. No. 9.2e-07;  
Matches 73; Conservative 30; Mismatches 98; Indels 78; Gaps 10;  
QY 1 MIFIAVGNTRTLAHTHDGVHFDVSVATSLPPTTEILQQPGLTWLSAPNREPVALGGVVP 60

Db 13 LLLIDAGNTRIKWAWT-----AADVAPPA--VAPGGTPWQHAGARPHDQLAELV- 59  
QY 61 AALAAWREALATAEVREPDGFFRRVAP----- 88  
Db 60 ---EDWRDCHAGAGMAPPDVWISVAGPALRDALCARIARVFDGARLRIVASEAAAAAGLR 116  
QY 89 HDYHPPEISLGFDRRCCLLAAAMDYPGQDSIVDMGTAITIDLLA-GSHFRGGRILPGIAM 147  
Db 117 NGYRDPALQGTDRWVGAVGARHAWPDTALLLVTAGTATTLDIVAPDGRFAGGLILPGLTL 176  
QY 148 SLRGLHEGTALLPEVVL-----NAPAEW--LGNDTSNAIQAGVHILFADALRGAIT 196  
Db 177 MMRALSRNTAQLPEIDIGYLAARDDAQAPADVPWSWADNTQDAIALGCV-----TAQAGAIA 232  
QY 197 DF-----RQYSPQARILITGGDAERWQPGIAGSLYQPHL 230  
Db 233 QTWQALQAQYPCPYRCVLSGG-----ARAALAPHL 262

RESULT 12

Q9KGH5 PRELIMINARY; PRT; 254 AA.  
AC Q9KGH5;  
DT 01-OCT-2000 (TReMBLrel. 15, Created)  
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Hypothetical protein BH0086.  
GN BH0086.  
OS Bacillus halodurans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Bacillaceae; Bacillus.  
OX NCBI\_TaxID=86665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C-125 / JCM 9153;  
RX MEDLINE=20512582; PubMed=11058132;  
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,  
RA Horikoshi K.;  
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
halodurans and genomic sequence comparison with Bacillus subtilis.";  
RL Nucleic Acids Res. 28:4317-4331(2000).  
DR EMBL; AP001507; BAB03805.1; -  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMS; TIGR00671; baf; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 254 AA; 27907 MW; 0E3B557BA7DAC176 CRC64;

Query Match 14.2%; Score 179.5; DB 16; Length 254;  
Best Local Similarity 26.1%; Pred. No. 8.4e-07;  
Matches 61; Conservative 47; Mismatches 97; Indels 29; Gaps 9;  
QY 1 MIFIAVGNTRTLAHTDGVHFDVSV-----ATSLPPT-----ILQQPLTWLS 46  
Db 2 ILVIDVGNNTVL-----GVYQDETLLVHHWRLATSRQKTEDEYAMTVRSFLDHAGLQF-- 54  
QY 47 APNREPVALGGVVPAAALAAWREALATAEVREP---DPGFFRRVAVPHDYHPPESLGFDRRC 103  
Db 55 -QDIDGIVISSVWPPPMFSLQMKCKYFHVTPMIIGPG-IKTGLNIKYDNPKEVGAD-RI 111  
QY 104 CLLAAAMDYPGQDSIVDMGTAITIDLL-AGGHFRGGRILPGIAMSRLGHEGTALLPEV 162  
Db 112 VNAAIAELYGYPVIVDFGTATTYCLINEKKQYAGGVIAPGIMISTEALYHRASKLPRI 171  
QY 163 VLNAPAEMLGNDTSNAIQAGVHILFADALRGAITDFR-QYSPQARILITGGDAE 215  
Db 172 EIAKPKVQVGTNTIDSMQSGIFYGVSVQDGVVKRMKAQAESEPKVIATGGLAK 225

RESULT 13

P74045

ID P74045 PRELIMINARY; PRT; 257 AA.  
AC P74045;  
DT 01-FEB-1997 (TReMBLrel. 02, Created)  
DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Hypothetical protein sir0812.  
GN SLR0812.  
OS Synechocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
OX NCBI\_TaxID=1148;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97061201; PubMed=8905231;  
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
RA Miyajima N., Hiroseawa M., Sugiyura M., Sasamoto S., Kimura T.,  
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,  
RA Shimpō S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,  
RA Tabata S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
Synechocystis sp. strain PCC6803. II. Sequence determination of the  
entire genome and assignment of potential protein-coding regions.";  
RL DNA Res. 3:109-136(1996).  
DR EMBL; D90911; BAA18120.1; -  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMS; TIGR00671; baf; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 257 AA; 27702 MW; 4291E51EE91FB4A2 CRC64;  
Query Match 13.8%; Score 174; DB 16; Length 257;  
Best Local Similarity 27.9%; Pred. No. 2.4e-06;  
Matches 60; Conservative 34; Mismatches 73; Indels 48; Gaps 8;  
QY 40 PGLTWLSAPN-----REPVALGGVVPAAALAAWREALATAEVREPDGFFRRVAP 88  
Db 40 PLQTWVTVDYNPKSAQLPVLGKVPMLASVPEQTEVWR-----VYQPKILTILKNLPL 92  
QY 89 HDYHPPEISLGFDRRCCLLAAAMDYPGQDSIVDMGTAITIDLLAGGHFR-----GGRILP 143  
Db 93 VNLYP--SFGIDRALAGLGTGLTY-GFPCLVVDGGTALTIT---GFDQDKLVGGAILP 145  
QY 144 GIAMSLRGLHEGTALLPEVVLNAPAEMLGN---DTSNAIQAGVHILFADALRGAITDFRQ 200  
Db 146 GIGLQLATIGDRLAALPKLEMDQLTELPDRWALDTPSAIFSGVYGVGLGALQSYIQDWOK 205  
QY 201 YSPQARILITGGDAERWQPGIAGSLYQPHLLLRGF 235  
Db 206 LFPGAAMVITGGDGK-----ILHGF 225

RESULT 14

Q8XHL5 PRELIMINARY; PRT; 259 AA.  
ID Q8XHL5  
AC Q8XHL5;  
DT 01-MAR-2002 (TReMBLrel. 20, Created)  
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Hypothetical protein CPE2468.  
GN CPE2468.  
OS Clostridium perfringens.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
OC Clostridiales; Clostridiaceae; Clostridium.  
OX NCBI\_TaxID=1502;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=13 / TYPE A;  
RX PubMed=11792842;  
RA Shlimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,  
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;  
RT "Complete genome sequence of Clostridium perfringens, an anaerobic  
flesh-eater.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).  
DR EMBL; AP003194; BAB82174.1; -





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

.OM protein - protein search, using sw model

Run on: June 24, 2003, 21:49:47 ; Search time 9.77804 Seconds  
(without alignments)  
725.188 Million cell updates/sec

Title: US-09-813-453A-63  
 Perfect score: 1264  
 Sequence: 1 MIFIAGVNTRTLAAHTHDGV.....AGSLYQPHLLLRGFLYMWIRG 241

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters:	262574
--	--------

```
Minimum DB seq length: 0
Maximum DB seq length: 2
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```
Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	98.5	7.8	6095	4	US-09-144-085-2	Sequence 2, Appli
2	94.5	7.5	5215	4	US-09-105-537-2	Sequence 2, Appli
3	92	7.3	5087	4	US-09-144-085-1	Sequence 1, Appli
4	86.5	6.8	3724	2	US-08-804-227C-10	Sequence 10, Appli
5	86.5	6.8	3724	2	US-08-804-198-4	Sequence 4, Appli
6	86	6.8	1077	4	US-09-412-210-1	Sequence 1, Appli
7	86	6.8	1864	2	US-08-804-227C-3	Sequence 3, Appli
8	85.5	6.8	4545	2	US-08-804-227C-14	Sequence 14, Appli
9	85.5	6.8	4550	2	US-08-804-227C-8	Sequence 8, Appli
10	85.5	6.8	4550	2	US-08-804-198-2	Sequence 2, Appli
11	82	6.5	654	4	US-09-620-412C-341	Sequence 341, App
12	82	6.5	1752	4	US-09-556-877-180	Sequence 180, App
13	82	6.5	1752	4	US-09-620-412C-180	Sequence 180, App
14	80	6.3	3567	2	US-07-642-734C-4	Sequence 4, Appli
15	80	6.3	3567	3	US-08-439-009A-4	Sequence 4, Appli
16	78.5	6.2	1399	4	US-08-462-467B-14	Sequence 14, Appli
17	78	6.2	1958	1	US-07-945-283-2	Sequence 2, Appli
18	77.5	6.1	3816	4	US-09-428-517-3	Sequence 3, Appli
19	77	6.1	485	4	US-09-384-212-2	Sequence 2, Appli
20	77	6.1	1317	3	US-09-083-521-7	Sequence 7, Appli
21	77	6.1	1891	2	US-08-804-227C-12	Sequence 12, Appli
22	77	6.1	1891	2	US-08-804-198-6	Sequence 6, Appli
23	76	6.0	396	4	US-09-046-992-4	Sequence 4, Appli
24	76	6.0	420	1	US-08-391-259-2	Sequence 2, Appli
25	76	6.0	420	1	US-08-391-259-7	Sequence 7, Appli
26	76	6.0	420	1	US-08-391-259-10	Sequence 10, Appli
27	76	6.0	420	1	US-08-391-259-11	Sequence 11, Appli

## ALIGNMENTS

```

RESULT 1
US-09-144-085-2
; Sequence 2, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
; APPLICANT: Gustafsson, Claes
; APPLICANT: Betlach, Mary C.
; APPLICANT: Ashley, Gary
; APPLICANT: Julien, Bryan
; APPLICANT: Ziermann, Rainer
; TITLE OF INVENTION: SORANGIUM POLY
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 30062-20020.20
; CURRENT APPLICATION NUMBER: US/09/
; CURRENT FILING DATE: 1998-08-31
; EARLIER APPLICATION NUMBER: 09/010
; EARLIER FILING DATE: 1998-01-22
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 6095
; TYPE: PRT
; ORGANISM: Sorangium cellulosum.
US-09-144-085-2

```

	Query Match	7.8%;	Score 98.5;	DB 4;	Length 6095;
	Best Local Similarity	26.0%;	Pred. NO. 0.73;		
	Matches	58;	Conservative	21;	Mismatches 79; Indels 65; Gaps 10;
QY	11	TLLAHTDGVHFDVS	VATSLPPTTEILQQGLTWLSAPNREPVALGGVVPAALAAWREAL	70	
		:     :	:     :     :     :     :		
Dd	1293	TALGQSQSGV--ASLSHEAT	RQALSLLQ-----AWLSEPRLD	AVELVWVTRGAVGAAPDDA	1346
			:		
QY	71	ATAEVREPDPGFFERRAVPHDYHPPE	SLGDFDRRCCLLAAMDYPGQDSIVIDMGT-AITID	129	
			:     :     :     :     :		
Dd	1347	VQDLARAPLWGLVR-----	-----AARSEHPERRRLRIDVGTEPVDAG	1384	
			:		
QY	130	LLAGGHFRGGRILPGIAMSLRGLHEGTALLPEVVNLNAPAEMLGNDTSNAIQAGVIHLFAD	189		
			:     :     :     :     :		
Dd	1385	LIA-----RAL--ATAEPFELARGGA-----	-----ALAARLVRAQAA	1416	
			:		
QY	190	A---LRGALTDFEROYSPOARILITGGDAERWQP	GIAGSLYQPH	229	
			:     :     :     :     :		
Dd	1417	AEELTRGA----RELDPAGTVLVTGGTGELGO-AIAAHLVRAH	1454		

RESULT 2  
US-09-105-537-2  
; Sequence 2, Application US/09105537A  
; Patent No. 6265202





Db 3206 ---AGPDALGRLLRELMDFLAAGVLHPLPVVTHDVRRRAADALR-TISQARHTGKLVLTMP 3261  
QY 201 --YSPQARILITGGDAERWQPGIAGSLYQPHLLR 233  
Db 3262 PAWHPYGTVLVTGG-----TGALGSRIRARHLASR 3290

RESULT 5  
US-08-804-198-4  
; Sequence 4, Application US/08804198  
; Patent No. 5945320  
; GENERAL INFORMATION:  
; APPLICANT: Burgett, Stanley G.  
; APPLICANT: Kuhstoss, Stuart A.  
; APPLICANT: Rao, Nagaraja R.  
; APPLICANT: Richardson, Mark A.  
; APPLICANT: Rostock, Paul R., Jr.  
; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PAUL R. CANTRELL 1138  
; STREET: LILLY CORPORATE CENTER  
; CITY: INDIANAPOLIS  
; STATE: IN  
; COUNTRY: USA  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Macintosh  
; OPERATING SYSTEM: Macintosh 7.0  
; SOFTWARE: Microsoft Word 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/804,198  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CANTRELL, PAUL R.  
; REGISTRATION NUMBER: 36,470  
; REFERENCE/DOCKET NUMBER: P9113  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-3885  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3724 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide

US-08-804-198-4  
Query Match 6.8%; Score 86.5; DB 2; Length 3724;  
Best Local Similarity 25.8%; Pred. No. 7.5;  
Matches 71; Conservative 21; Mismatches 102; Indels 81; Gaps 14;  
QY 12 LLAHTDGVHFDVSVATSLPPTTEILQQPGLTWLSA--PNREPV---ALGGVVPAALAAW 66  
Db 3044 LLAPVPDGW---SYAQAAASVPAVFLSAYIGLVTLAGLRPGERVLVHAAAGVGMAAVQIA 3100  
QY 67 R---EALATAEVREPDPG---FFRAVPHDYHPPESLGFDRRCCLLAAAMDYPGQDSIV 119  
Db 3101 RHLGAEVLATA-----SPGKWDALRAMGITDDHLASSRTLDEAFTGA--DGTSRADV 3153  
QY 120 IDMGTAITIDLLAGHFRGGRIL-----PGIAMSRLRGLHEGTALLPEV 162  
Db 3154 LNSLTKEFVDASLGLLRPGGRFLELGTVDVDPERIAAEHPGVRYRAFDLNE----- 3205  
QY 163 VLNAPAEMLG---NDTSNAIQAGVIHLF-----ADALRGAITDPRO----- 200  
Db 3206 ---AGPDALGRLLRELMDFLAAGVLHPLPVVTHDVRRRAADALR-TISQARHTGKLVLTMP 3261  
QY 201 --YSPQARILITGGDAERWQPGIAGSLYQPHLLR 233  
Db 3262 PAWHPYGTVLVTGG-----TGALGSRIRARHLASR 3290

RESULT 6  
US-09-412-210-1  
; Sequence 1, Application US/09412210  
; Patent No. 6403358  
; GENERAL INFORMATION:  
; APPLICANT: Kapeller-Libermann, Rosana  
; TITLE OF INVENTION: 21529, A NOVEL ADENYLATE CYCLASE  
; FILE REFERENCE: 5800-47  
; CURRENT APPLICATION NUMBER: US/09/412,210  
; CURRENT FILING DATE: 1999-10-05  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 1077  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-412-210-1

Query Match 6.8%; Score 86; DB 4; Length 1077;  
Best Local Similarity 25.9%; Pred. No. 1.4;  
Matches 58; Conservative 23; Mismatches 66; Indels 106; Gaps 16;  
QY 32 PPTTEILQQPGLTWLSAPNREP--VALGGVVPAALAAWREALATAEVRE--PDPGFFERRAV 87  
Db 10 PPSEDLFE--TYISLSQQYPLLLLLLGLVLCALALL-AVAVASGRELTSDPSFLTTL 66  
QY 88 PHDYHPPESLGFDRRCCLLAAAMDYPGQDSIVIDMGT-----AITIDLLAG 133  
Db 67 -----CAL-----GGFSLGLGLASREORLQRWTRPLSLGLVWVALLAL 103  
QY 134 GH---FRGG-----RIIP-----GIAMSLR-----GLHEG-- 155  
Db 104 GHAFLETTGGVVSANDQVSVELFVIFTAYAMPLGMRDAAVAGLASSLSHLLVLGLYLPQ 163  
QY 155 ----TALLPEVVLNAPAEMLGNDTSNAIQAGVIH--LFADALRGAITDFRQYSPQARILI 209  
Db 164 PDSRPALLPQLAANAVLFCGN-----VAGVYHKALMERALRATFREALSLSLHRRRL- 216  
QY 210 TGGDAERWQPGIAGSLYQPHLLL 232  
Db 217 ---DTEK-----KHQEHLLL 228

RESULT 7  
US-08-804-227C-3  
; Sequence 3, Application US/08804227C  
; Patent No. 5876991  
; GENERAL INFORMATION:  
; APPLICANT: DeHoff, Bradley S.  
; APPLICANT: Kuhstoss, Stuart A.  
; APPLICANT: Rostock, Paul R., Jr.  
; APPLICANT: Sutton, Kimberly L.  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: THOMAS G. PLANT 1501  
; STREET: LILLY CORPORATE CENTER  
; CITY: INDIANAPOLIS  
; STATE: IN  
; COUNTRY: USA  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: ASCII(DOS) Text only  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/804,227C  
; FILING DATE: February 21, 1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:

NAME: Plant, Thomas, G.  
REGISTRATION NUMBER: 35,784  
REFERENCE/DOCKET NUMBER: X-8231  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-2459  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1864 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-804-227C-3

Query Match 6.8%; Score 86; DB 2; Length 1864;  
Best Local Similarity 27.8%; Pred. No. 3.1;  
Matches 45; Conservative 14; Mismatches 43; Indels 60; Gaps 9;  
Qy 24 SVSVATSLPPTTEILQQPGLTWLSAPNREPVALGGVVPAAALAAWREALATAEVRPDGFF 83  
Db 920 SLAVAAEL-PTYAFORTHY-WLDAP-----AAPALPAGLD-----DAG-- 956  
Y 84 RRAVPHDYHPPEGLFDRRCCLAAAMDYP-GQDSIVIDMGTAITIDLLAGGHFRGGRIL 142  
Db 957 -----HP-----LLSAALDLPGGRGTVTGALSATLPWAADHSVHGRTVL 997  
Qy 143 PGIAMSLRGLH-----EGTALLPE-----VVLNAP 167  
Db 998 PGTALLDLALHAAPRVGELTFEAPLVLPEDGEVRLRVVLAEP 1039

RESULT 8  
US-08-804-227C-14  
; Sequence 14, Application US/08804227C  
; Patent No. 5876991  
; GENERAL INFORMATION:  
; APPLICANT: DeHoff, Bradley S.  
; APPLICANT: Kuhstoss, Stuart A.  
; APPLICANT: Rostock, Paul R., Jr.  
; APPLICANT: Sutton, Kimberly L.  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: THOMAS G. PLANT 1501  
; STREET: LILLY CORPORATE CENTER  
; CITY: INDIANAPOLIS  
; STATE: IN  
; COUNTRY: USA  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: ASCII(DOS) Text only  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/804,227C  
; FILING DATE: February 21, 1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plant, Thomas, G.  
; REGISTRATION NUMBER: 35,784  
; REFERENCE/DOCKET NUMBER: X-8231  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-2459  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4545 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-804-227C-14

Query Match 6.8%; Score 85.5; DB 2; Length 4545;  
Best Local Similarity 21.9%; Pred. No. 13;

Matches 63; Conservative 27; Mismatches 105; Indels 93; Gaps 14;  
Qy 11 TLLAHTHDGVHF-DSVSVATSLPPTTEILQQPGLTWLSAPNREPVALGGV---PAALAA- 65  
Db 2604 SLASHLHDELFGPDSEAEPAAPTPVM-----ADEREPIAIVGMACRYPGGVASP 2654  
Qy 66 ---WREALATAEVRPDP-----GFFRRAVPHDYHPPE 96  
Db 2655 DDLWDLVAGDGHLSPPFADRGWDVEGLYDPEPGVPGKSYVREGGFLRSAAEFD---AEF 2711  
Qy 97 LGFDRRCCLAAAMDYPGQDSIVIDMGTAITIDLLAGGHFRGGR---ILPGI----- 145  
Db 2712 FGISPR---EATAMD-PQORLLLETSWREALERAGIVPDSLRTGTGVFSGISQQDYATQL 2767  
Qy 146 ---AMSLRGLHEGTALLPEVVLNAPAEMLG-----NDTSNAIQAGVIHLFADALR---- 192  
Db 2768 GDAADTYGGHVLGTGLSGVISGRVAYALGLEGPALTVDTACSSSLVALHLAVQSLRRGEC 2827  
Qy 193 -----GAI-----TDFROYSPQ-----ARILTTGGDAERWQPGI 221  
Db 2828 DLALAGGVTVMATPTVFVEFSRQRLAADRCKAFAGADGTAWAEGV 2875

RESULT 9  
US-08-804-227C-8  
; Sequence 8, Application US/08804227C  
; Patent No. 5876991  
; GENERAL INFORMATION:  
; APPLICANT: DeHoff, Bradley S.  
; APPLICANT: Kuhstoss, Stuart A.  
; APPLICANT: Rostock, Paul R., Jr.  
; APPLICANT: Sutton, Kimberly L.  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: THOMAS G. PLANT 1501  
; STREET: LILLY CORPORATE CENTER  
; CITY: INDIANAPOLIS  
; STATE: IN  
; COUNTRY: USA  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: ASCII(DOS) Text only  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/804,227C  
; FILING DATE: February 21, 1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plant, Thomas, G.  
; REGISTRATION NUMBER: 35,784  
; REFERENCE/DOCKET NUMBER: X-8231  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-2459  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4550 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
US-08-804-227C-8

Query Match 6.8%; Score 85.5; DB 2; Length 4550;  
Best Local Similarity 21.9%; Pred. No. 13;  
Matches 63; Conservative 27; Mismatches 105; Indels 93; Gaps 14;  
Qy 11 TLLAHTHDGVHF-DSVSVATSLPPTTEILQQPGLTWLSAPNREPVALGGV---PAALAA- 65  
Db 2609 SLASHLHDELFGPDSEAEPAAPTPVM-----ADEREPIAIVGMACRYPGGVASP 2659  
Qy 66 ---WREALATAEVRPDP-----GFFRRAVPHDYHPPE 96

Db 2660 DDLWDLVAGDGHGTLSPFPADRGWDVEGLYDPEPGVPGKSYVREGGFLRSAAEFD---AEF 2716  
QY 97 LGFDRRCCLLAAMDYPGQDSIVIDMGTAITIDLLAGGHFRGGR--ILPGI----- 145  
Db 2717 FGISPR---EATAMD-PQORLLLETSWALERAGIVPDSLGRTRTGVSFSGISQQDYATQL 2772  
QY 146 --AMSLRGLHEGTALLPEVVLNAPAEMLG-----NDTSNAIQAGVIHLFADALR--- 192  
Db 2773 GDAADTYGGHVLGTGLSGVISGRVAYALGLEGPALTVDTACSSSLVALHLAVQSLRRGEC 2832  
QY 193 -----GAI-----TDFRQYSPQ-----ARILITGGDAERWQPGI 221  
Db 2833 DLALAGGVTVMATPTVFEVFSRQRLAADGRCKAFAEGADGTAWAEGV 2880

RESULT 10  
US-08-804-198-2  
Sequence 2, Application US/08804198  
Patent No. 5945320  
GENERAL INFORMATION:  
APPLICANT: Burgett, Stanley G.  
APPLICANT: Kuhstoss, Stuart A.  
APPLICANT: Rao, Nagaraja R.  
APPLICANT: Richardson, Mark A.  
APPLICANT: Rosteck, Paul R., Jr.  
TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PAUL R. CANTRELL 1138  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Macintosh  
OPERATING SYSTEM: Macintosh 7.0  
SOFTWARE: Microsoft Word 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,198  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: CANTRELL, PAUL R.  
REGISTRATION NUMBER: 36,470  
REFERENCE/DOCKET NUMBER: P9113  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-3885  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4550 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-804-198-2

Query Match 6.8%; Score 85.5; DB 2; Length 4550;  
Best Local Similarity 21.9%; Pred. No. 13;  
Matches 63; Conservative 27; Mismatches 105; Indels 93; Gaps 14;  
QY 11 TLLAHTHDGVHF-DSVSVATSLPPTTEILQQPGLTWLSAPNREPVALGGVV---PAALAA- 65  
Db 2609 SLASHLHDELFGPDSEAEPAAPTPVM-----ADERPIAVGMACRYPGGVAS 2659  
QY 66 ---WREALATAEVRPDP-----GFFERRAVPHDYHPPES 96  
Db 2660 DDLWDLVAGDGHGTLSPFPADRGWDVEGLYDPEPGVPGKSYVREGGFLRSAAEFD---AEF 2716  
QY 97 LGFDRRCCLLAAMDYPGQDSIVIDMGTAITIDLLAGGHFRGGR--ILPGI----- 145  
Db 2717 FGISPR---EATAMD-PQORLLLETSWALERAGIVPDSLGRTRTGVSFSGISQQDYATQL 2772

QY 146 --AMSLRGLHEGTALLPEVVLNAPAEMLG-----NDTSNAIQAGVIHLFADALR--- 192  
Db 2773 GDAADTYGGHVLGTGLSGVISGRVAYALGLEGPALTVDTACSSSLVALHLAVQSLRRGEC 2832  
QY 193 -----GAI-----TDFRQYSPQ-----ARILITGGDAERWQPGI 221  
Db 2833 DLALAGGVTVMATPTVFEVFSRQRLAADGRCKAFAEGADGTAWAEGV 2880  
RESULT 11  
US-09-620-412C-341  
Sequence 341, Application US/09620412C  
Patent No. 6448234  
GENERAL INFORMATION:  
APPLICANT: Steven P. Fling  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION  
FILE REFERENCE: 210121.469C7  
CURRENT APPLICATION NUMBER: US/09/620,412C  
CURRENT FILING DATE: 2000-07-20  
NUMBER OF SEQ ID NOS: 363  
SOFTWARE: FastSEQ for Windows Version 3.0/4.0  
SEQ ID NO 341  
LENGTH: 654  
TYPE: PRT  
ORGANISM: Chlamydia trachomatis  
US-09-620-412C-341

Query Match 6.5%; Score 82; DB 4; Length 654;  
Best Local Similarity 21.2%; Pred. No. 1.9;  
Matches 62; Conservative 43; Mismatches 114; Indels 74; Gaps 13;  
QY 4 IAVGNTRTLLAHTHDGVHFDVSVVATSLPPTTEILQQPGLTWLSAPNREPVALGGV----- 58  
Db 176 VLSVTENLVGKEGGLHAKTVNISN-----LKSGFSF--SNKANSSSTGVATTAS 225  
QY 59 VPAALAAWREALATAEVRPDPGFFRRVPHDYHPPESLGFDRC---CLLA---AAMDY 112  
Db 226 AAAAAAASLQAAAAAAPSPATPTYSGVVGAIY-GEKVTF-QCSGTCQFSGNQAIIDNN 283  
QY 113 PGQDSIVIDMG-----TAITIDLLAGGH---FRGGRILPGIAMSRLRGLHEGTALLPEVVL 164  
Db 284 PSQSSLNQGGAIYAKTSLSGSSDAGTSYIFSGNSVSTGKSQTTGQIAGGAIYSPTVTL 343  
QY 165 NAPAEMLGNDTS-----NAIQ-----AGV----- 183  
Db 344 NCPATFSNNTASIAIATPKTSSDSSGNSIKDITIGGAIAGTATLGSVSRFSGNTADLGAA 403  
QY 184 IHLFADALRGAITDFRQYSPQARILITGGD--AERWQPGIAGSLYQPHLLLRG 234  
Db 404 IGTLANANTPSATSGSQNSITEKITLNGSFIFERNQANKRGAIYSPSVSIKG 456

RESULT 12  
US-09-556-877-180  
Sequence 180, Application US/09556877  
Patent No. 6432916  
GENERAL INFORMATION:  
APPLICANT: Probst, Peter  
APPLICANT: Bhatia, Ajay  
APPLICANT: Skeiky, Yasir  
APPLICANT: Fling, Steve  
APPLICANT: Maisonneuve, Jeff  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION  
FILE REFERENCE: 210121.469C5  
CURRENT APPLICATION NUMBER: US/09/556,877  
CURRENT FILING DATE: 2000-04-19  
NUMBER OF SEQ ID NOS: 305  
SOFTWARE: FastSEQ for Windows Version 3.0/4.0  
SEQ ID NO 180  
LENGTH: 1752

TYPE: PRT  
ORGANISM: Chlamydia  
US-09-556-877-180

Query Match 6.5%; Score 82; DB 4; Length 1752;  
Best Local Similarity 21.2%; Pred. No. 7.8;  
Matches 62; Conservative 43; Mismatches 114; Indels 74; Gaps 13;

QY 4 IAVGNTRLLAHTHDGVHFDVSVAATSLPPTTEILQQPGLTWLSAPNREPVALGGV----- 58  
Db VLSVTENLVGKEGGLHAKTVNISN-----LKSQFSF--SNNKANSSSTGVATTAS 809  
QY 59 VPAALAAWREALATAEVREPDGFFRRAVPHDYHPPESLGFDRRC---CLLA---AAMDY 112  
Db APAAAAASLQAAAAAAPSPATPTYSGVVGGAIIY-GEKVTF--QCSGTCQFSGNQADNN 867  
QY 113 PGQDSIVIDMG-----TAITIDLLAGGH---FRGGRILPGIAMSRLRGLHEGTALLPEVVL 164  
Db PSQSSLNQVGGGAIYAKTSLSIGSSDAGTSYIFSGNSVSTGKSQTTGQIAGGAIYSPVTTL 927  
QY 165 NAPAEMLGNDTS-----NAIQ-----AGV----- 183  
Q 928 NCPATFSNNTASIAATPKTSSDSSGNSIKDTIGGAIAGTAITLSGVSRFSGNTADLGAA 987  
QY 184 IHLFADALRGAITDFROYSPQARILITGGD--AERWQPGIAGSLYQPHLLLRG 234  
Db IGTLANANTPSATSGSQNSITEKITLENGSIIFERNQANKRGAIYSPSVSIKG 1040

RESULT 13

US-09-620-412C-180  
Sequence 180, Application US/09620412C  
Patent No. 6448234

GENERAL INFORMATION:

APPLICANT: Steven P. Fling  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
DIAGNOSIS OF CHLAMYDIAL INFECTION

FILE REFERENCE: 210121.469C7  
CURRENT APPLICATION NUMBER: US/09/620,412C  
CURRENT FILING DATE: 2000-07-20  
NUMBER OF SEQ ID NOS: 363  
SOFTWARE: FastSeq for Windows Version 3.0/4.0

SEQ ID NO 180  
LENGTH: 1752

TYPE: PRT  
ORGANISM: Chlamydia  
US-09-620-412C-180

Query Match 6.5%; Score 82; DB 4; Length 1752;  
Best Local Similarity 21.2%; Pred. No. 7.8;  
Matches 62; Conservative 43; Mismatches 114; Indels 74; Gaps 13;

QY 4 IAVGNTRLLAHTHDGVHFDVSVAATSLPPTTEILQQPGLTWLSAPNREPVALGGV----- 58  
Db VLSVTENLVGKEGGLHAKTVNISN-----LKSQFSF--SNNKANSSSTGVATTAS 809  
QY 59 VPAALAAWREALATAEVREPDGFFRRAVPHDYHPPESLGFDRRC---CLLA---AAMDY 112  
Db APAAAAASLQAAAAAAPSPATPTYSGVVGGAIIY-GEKVTF--QCSGTCQFSGNQADNN 867  
QY 113 PGQDSIVIDMG-----TAITIDLLAGGH---FRGGRILPGIAMSRLRGLHEGTALLPEVVL 164  
Db PSQSSLNQVGGGAIYAKTSLSIGSSDAGTSYIFSGNSVSTGKSQTTGQIAGGAIYSPVTTL 927  
QY 165 NAPAEMLGNDTS-----NAIQ-----AGV----- 183  
Db NCPATFSNNTASIAATPKTSSDSSGNSIKDTIGGAIAGTAITLSGVSRFSGNTADLGAA 987  
QY 184 IHLFADALRGAITDFROYSPQARILITGGD--AERWQPGIAGSLYQPHLLLRG 234  
Db IGTLANANTPSATSGSQNSITEKITLENGSIIFERNQANKRGAIYSPSVSIKG 1040

RESULT 14

US-07-642-734C-4  
Sequence 4, Application US/07642734C  
Patent No. 5824513

GENERAL INFORMATION:

APPLICANT: Katz, L  
APPLICANT: Donadio, S  
APPLICANT: Mcalpine, J B  
TITLE OF INVENTION: Recombinant DNA Method for Producing  
Erythromycin Analogs  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Edward H. Gorman  
STREET: Abbott Laboratories D377/AP6D-2 One Abbott  
CITY: Abbott Park  
STATE: IL  
COUNTRY: US  
ZIP: 60064-3500

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/642,734C  
FILING DATE: 17-JAN-91  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Dancakers, Andreas M  
REGISTRATION NUMBER: 32652  
REFERENCE/DOCKET NUMBER: 4952.US.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 708-937-9396  
TELEFAX: 708-938-2623  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3567 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-07-642-734C-4

Query Match 6.3%; Score 80; DB 2; Length 3567;  
Best Local Similarity 23.9%; Pred. No. 37;  
Matches 58; Conservative 13; Mismatches 60; Indels 112; Gaps 11;

QY 54 ALGGVVPALA-----AWREALATAEVREPDGFFRRAVPHDYHPPES--LGFDRRC--- 104  
Db AAGVGMAAVALARRAGAEVLATAGPAK--HGTLRALGLDDEHIASSRETGFARKFRERT 3023  
QY 105 -----LLAAMDYPGQDSIVIDMGTAITIDLLAGGHFRG----- 138  
Db GGRGVDVVLNSLTGELLDESADLLAEDGVFVEMGKT---DLRDAGDFGRVAPFDLGEAG 3080  
QY 139 -----GRI-----LPGIAMSL-----RGLHEGTALLPEVVLNAPA 168  
Db DDRLGEILREVVGLLGAGELDRLPVSAWELGSAPAAQHMSRGRHVG-----KLVLTQPA 3135  
QY 169 EMLGNDTSNAIQAGVIHLFADALRGAITDFROYSPQARILITGGDAERWQPGIAGSLYQP 228  
Db 3136 PV-----DPDGTVLITGG-----TGTLGRLLAR 3158  
QY 229 HLL 231  
Db 3159 HL 3161

RESULT 15

US-08-439-009A-4  
Sequence 4, Application US/08439009A  
Patent No. 6004787

GENERAL INFORMATION:





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

-OM protein - protein search, using sw model

Run on: June 24, 2003, 22:03:27 ; Search time 9.94238 seconds  
(without alignments)  
2330.267 Million cell updates/sec

Title: US-09-813-453A-63  
Perfect score: 1264  
Sequence: 1 MFIAGVGNTRTLAHTHDGV.....AGSLYQPHLLLRGFYLWIRG 241

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

sarched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	229.5	18.2	246	2 D72320	conserved hypothet
2	228	18.0	261	2 B87489	transcription acti
3	215	17.0	262	2 E75516	conserved hypothet
4	190	15.0	276	2 AI2292	hypothetical prote
5	185	14.6	259	2 AF1102	conserved hypothet
6	184	14.6	265	2 T36391	hypothetical prote
7	181	14.3	233	2 S66100	conserved hypothet
8	181	14.3	259	2 AF1464	conserved hypothet
9	179.5	14.2	254	2 F83660	hypothetical prote
10	174	13.8	257	2 S75559	hypothetical prote
11	172	13.6	274	2 H86937	conserved hypothet
12	170.5	13.5	273	2 D71326	conserved hypothet
13	164.5	13.0	229	2 E70465	hypothetical prote
14	159	12.6	272	2 A70955	hypothetical prote
15	158.5	12.5	273	2 E97293	probable transcrip
16	154.5	12.2	267	2 I40327	baf protein - Bord
17	144.5	11.4	223	2 F64627	hypothetical prote
18	144.5	11.4	248	2 H83111	hypothetical prote
19	143.5	11.4	223	2 G71887	hypothetical prote
20	143.5	11.4	242	2 A82637	conserved hypothet
21	136	10.8	262	2 F70165	conserved hypothet
22	124	9.8	209	2 H81382	hypothetical prote
23	115.5	9.1	592	2 B81009	Bira protein/Bvg a
24	112.5	8.9	592	2 H82031	probable biotin-[a
25	96.5	7.6	558	2 S27199	cephalosporin acyl
26	95.5	7.6	430	2 T44851	molybdopterin co-f
27	95	7.5	2154	2 F83068	hypothetical prote
28	94.5	7.5	548	2 C75499	cytidine/deoxycyti
29	94.5	7.5	807	1 D75614	xanthine dehydroge

30	94.5	7.5	1827	2 B70984	probable polyketid
31	93	7.4	335	2 T34837	probable transfera
32	92	7.3	282	2 T36832	probable rRNA meth
33	92	7.3	10223	2 T30225	polyketide synthas
34	91.5	7.2	583	2 A70729	hypothetical prote
35	91	7.2	488	2 S27652	probable aldehyde
36	90.5	7.2	429	2 G87636	conserved hypothet
37	90.5	7.2	1860	2 T17485	peptide synthetase
38	90	7.1	314	2 T35241	hypothetical prote
39	90	7.1	423	2 S73020	hypothetical prote
40	90	7.1	558	1 B28392	penicillin amidase
41	90	7.1	693	2 C87575	sensor histidine k
42	89.5	7.1	515	2 AD1048	conserved hypothet
43	89.5	7.1	980	2 S71090	peroxisome biogene
44	89	7.0	469	2 AH3271	3-isopropylmalate
45	89	7.0	474	2 AD0083	probable cell divi

ALIGNMENTS

RESULT 1

D72320  
conserved hypothetical protein - Thermotoga maritima (strain MSB8)  
C;Species: Thermotoga maritima  
C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C;Accession: D72320  
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; H. Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, C.M.  
Nature 399, 323-329, 1999  
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome A;Reference number: A72200; MUID:99287316; PMID:10360571  
A;Accession: D72320  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-246 <ARN>  
A;Cross-references: GB:AE001754; GB:AE000512; NID:g4981417; PIDN:AAD35964.1; PID:g4; A;Experimental source: strain MSB8  
C;Genetics:  
A;Gene: TM0883  
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 18.2%; Score 229.5; DB 2; Length 246;  
Best Local Similarity 28.2%; Pred. No. 8.7e-13;  
Matches 71; Conservative 40; Mismatches 110; Indels 31; Gaps 5;

QY	2	IFIAVGNTRTLAHTHDGVHDSVSVATSLPPTTEILQQPGL-TWLSAPNREPVALG--GV	58
	:	:      : :     :    :    :    :    :    :    :    :    :    :	:
Db	3	LLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDELSHLPLLLGDAMREIKGIGVASV	62
	:	:      : :     :    :    :    :    :    :    :    :    :    :	:
QY	59	VPAALAA-----WREALATAEVREPPDPGFFRRRAVPHDYHPPESLGFDRRC	104
	:	:    :    :    :    :    :    :    :    :    :    :    :	:
Db	63	VPTONTVIERFSQKYPHISPIWVKAK-----NGCVKWNKNPSEVGADRVAN	109
	:	:    :    :    :    :    :    :    :    :    :    :    :	:
QY	105	LLAAAMDYPGQDSIVIDMGTAITIDLLAGGHFRGGRILPGIAMSRLRGLHEGTALLPEVVL	164
	:	:    :    :    :    :    :    :    :    :    :    :    :	:
Db	110	VVAFVKEY-GKNGIIDMGTTATVDLVVNGSYEGGAILPGFFMMVHSLFRGTAKPLPVEV	168
	:	:    :    :    :    :    :    :    :    :    :    :    :	:
QY	165	NAPAEMLGNDTSNAIQAGVIHLFADALRGAITDFRQYSPQARILITGGDAERWQPGIAGS	224
	:	:    :    :    :    :    :    :    :    :    :    :    :	:
Db	169	KPADFVVGKDTTEENIRLGVVNGSVYALEGIIIGRIKEVYGDLPVVLTTGGQSKIVKDMIKHE	228
	:	:    :    :    :    :    :    :    :    :    :    :    :	:
QY	225	LYQPHLLLRGFY	236
	:	:    :    :    :    :    :    :    :    :    :    :    :	:
Db	229	IFDEDLTIKGVY	240
	:	:    :    :    :    :    :    :    :    :    :    :    :	:

RESULT 2

B87489  
transcription activator, probable Baf family [imported] - Caulobacter crescentus  
C;Species: Caulobacter crescentus  
C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001















GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:51:40 ; Search time 4.84794 Seconds  
(without alignments)  
2061.866 Million cell updates/sec

Title: US-09-813-453A-63  
Perfect score: 1264  
Sequence: 1 MIFIAVGNTRTLATHDGV.....AGSLYQPHLLLRGFLWIRG 241

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

sarched: 112892 seqs, 41476328 residues 112892  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	181	14.3	233	1	YACB_BACSU	P37564 bacillus su
2	154.5	12.2	267	1	BAF_BORPE	Q45338 bordetella
3	96.5	7.6	557	1	PAC1_PSES	Q05053 pseudomonas
4	94	7.4	469	1	LEU2_RHIME	Q92176 rhizobium m
5	92	7.3	532	1	CKX1_ORISA	Q91de6 oryza sativ
6	92	7.3	690	1	VG42_BPMU	Q9tlv6 bacterioph
7	91.5	7.2	583	1	YP65_MYCTU	Q50733 mycobacteri
8	91	7.2	488	1	DHAL_PSESP	P33008 pseudomonas
9	90	7.1	557	1	PAC1_PSES3	P15557 pseudomonas
10	89	7.0	469	1	LEU2_BRUME	Q8Yjc9 brucella me
11	88.5	7.0	980	1	PEX6_HUMAN	Q13608 homo sapien
12	86	6.8	400	1	ARGD_MYCTU	P94990 mycobacteri
13	86	6.8	561	1	BCHD_RHOCA	P26175 rhodobacter
14	86	6.8	732	1	TRPE_AZOB	P50872 azospirillu
15	85	6.7	760	1	FTSH_MYCTU	P96942 mycobacteri
16	84.5	6.7	454	1	ATTY_HUMAN	P17735 homo sapien
17	84.5	6.7	1247	1	IRBP_HUMAN	P10745 homo sapien
18	84	6.6	286	1	AAC8_STRFR	P29809 streptomyce
19	84	6.6	1305	1	GAK_RAT	P97874 rattus norv
20	84	6.6	1733	1	VNUA_PRVKA	P33485 pseudorabie
21	83.5	6.6	525	1	P2BB_RAT	P20651 rattus norv
22	83	6.6	334	1	FEPD_ECOLI	P23876 escherichia
23	83	6.6	362	1	YD33_MYCLE	P53425 mycobacteri
24	83	6.6	629	1	YS50_MYCTU	Q05809 mycobacteri
25	82	6.5	1099	1	CARB_CAUCR	Q9a4d6 caulobacter
26	81	6.4	1754	1	PMPB_CHLTR	Q84418 chlamydia t
27	80	6.3	3567	1	ERY2_SACER	Q03132 saccharopol
28	79	6.2	469	1	LEU2_RHILO	Q98ef1 rhizobium l
29	78	6.2	247	1	SURE_THEMEA	P96112 thermotoga
30	78	6.2	691	1	DNLJ_MYCTU	O53261 mycobacteri
31	77.5	6.1	424	1	THC2_METTH	O27617 methanobact
32	77.5	6.1	515	1	YJEF_ECOLI	P31806 escherichia
33	77.5	6.1	700	1	PURL_HALNI	Q9hr49 halobacteri

RESULT 1  
YACB\_BACSU  
ID YACB\_BACSU STANDARD; PRT; 233 AA.  
AC P37564;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein yacB.  
GN YACB.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=96051385; PubMed=7584024;  
RA Ogasawara N., Nakai S., Yoshikawa H.;  
RT "Systematic sequencing of the 180 kilobase region of the Bacillus  
RT subtilis chromosome containing the replication origin.";  
RL DNA Res. 1:1-14(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
RA Guiseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
RA Lee S.N., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,  
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
RT subtilis.";  
RL Nature 390:249-256(1997).  
CC -1- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.

ALIGNMENTS

34	77	6.1	409	1	HUT1_RHILO	Q98311 rhizobium l
35	77	6.1	485	1	ANGT_HUMAN	P01019 homo sapien
36	77	6.1	637	1	MUTL_CAUCR	Q9rp66 caulobacter
37	77	6.1	1286	1	IRBP_BOVIN	P12661 bos taurus
38	75.5	6.1	1596	1	ACS2_ACEXY	Q59167 acetobacter
39	76	6.0	219	1	GPH_RHOCA	O33512 rhodobacter
40	76	6.0	390	1	YL28_STRCO	P40181 streptomyce
41	76	6.0	391	1	ARGD_CORGL	Q59282 corynebacte
42	76	6.0	453	1	MTEL_RAT	O55171 rattus norv
43	75.5	6.0	449	1	TIG_RALSO	Q8xyp8 ralstonia s
44	75.5	6.0	463	1	YV30_MYCTU	O07035 mycobacteri
45	75.5	6.0	597	1	WD42_DICDI	P54686 dictyosteli







DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Probable cytokinin oxidase precursor (EC 1.4.3.18) (CKO).  
GN P0512G09.9.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzeae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Nipponbare;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Catalyzes the oxidation of cytokinins, a family of  
CC N(6)-substituted adenine derivatives that are plant hormones,  
CC where the substituent is an isopentenyl group.  
CC -!- CATALYTIC ACTIVITY: N6-(3-methylbut-2-en-1-yl)adenine + H(2)O + O(2)  
CC - adenine + 3-methylbut-2-enal + H(2)O(2).  
CC -!- COFACTOR: FAD (By similarity).  
CC -!- SUBCELLULAR LOCATION: Extracellular (By similarity).  
CC -!- SIMILARITY: Belongs to the oxygen-dependent FAD-linked  
CC oxidoreductase family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AP002836; BAB07927.1; -  
CC EMBL; AP002816; BAB03420.1; -  
CC InterPro; IPR001575; Oxid\_FAD\_bind.  
CC Pfam; PF01565; FAD\_binding\_4; 1.  
CC PROSITE; PS00862; OX2\_COVAL\_FAD; 1.  
KW Oxidoreductase; Flavoprotein; FAD; Signal; Glycoprotein.  
FT SIGNAL 1 17 POTENTIAL.  
FT CHAIN 18 532 PROBABLE CYTOKININ OXIDASE.  
FT BINDING 105 105 FAD (COVALENT) (BY SIMILARITY).  
FT CARBOHYD 52 52 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 63 63 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 321 321 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 432 432 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 532 AA; 56034 MW; B849D646E3BB47E CRC64;  
Query Match 7.3%; Score 92; DB 1; Length 532;  
Best Local Similarity 24.3%; Pred. No. 1.3;  
Matches 46; Conservative 27; Mismatches 76; Indels 40; Gaps 9;  
QY 56 GGVVPAALAA--WREALATAEVRPDGFFRRVPHDYHPPESLGFDRCCLLAAMDY 112  
DB 24 GGGVPLAAAPLPFGDLAASGKLRDTP-----NATVPASMDFGNITAAALPAAVLF 74  
QY 113 PGQDSIVIDM-----GTAITIDLLAGGHRGGRILP--GIAMSLRGLHEGTALLPEV 162  
DB 75 PGSPGDVAELLRAAYAAPGRPFTVSRGRGHSTMGQALAAAGVVVHMOSMGGGA--PRI 132  
QY 163 VLNAPAEMLGNDTSNAIQAGVIHLFADALRGAITDFRQYSPQARI----LITGGDAERWQ 218  
DB 133 NVSADGAY-----VDAGGEQLWVDVLRALA--RGVAPRSWTDLHLTVGGTSL--N 180  
QY 219 PGIAGSLYQ 227  
DB 181 AGVSGQTYR 189  
RESULT 6  
VG42\_BPMU  
ID VG42\_BPMU  
AC Q9T1V6;  
Query Match 7.3%; Score 92; DB 1; Length 532;  
Best Local Similarity 24.3%; Pred. No. 1.3;  
Matches 46; Conservative 27; Mismatches 76; Indels 40; Gaps 9;  
QY 56 GGVVPAALAA--WREALATAEVRPDGFFRRVPHDYHPPESLGFDRCCLLAAMDY 112  
DB 24 GGGVPLAAAPLPFGDLAASGKLRDTP-----NATVPASMDFGNITAAALPAAVLF 74  
QY 113 PGQDSIVIDM-----GTAITIDLLAGGHRGGRILP--GIAMSLRGLHEGTALLPEV 162  
DB 75 PGSPGDVAELLRAAYAAPGRPFTVSRGRGHSTMGQALAAAGVVVHMOSMGGGA--PRI 132  
QY 163 VLNAPAEMLGNDTSNAIQAGVIHLFADALRGAITDFRQYSPQARI----LITGGDAERWQ 218  
DB 133 NVSADGAY-----VDAGGEQLWVDVLRALA--RGVAPRSWTDLHLTVGGTSL--N 180  
QY 219 PGIAGSLYQ 227  
DB 181 AGVSGQTYR 189  
RESULT 6  
VG42\_BPMU  
ID VG42\_BPMU  
AC Q9T1V6;  
Query Match 7.3%; Score 92; DB 1; Length 532;  
Best Local Similarity 24.3%; Pred. No. 1.3;  
Matches 46; Conservative 27; Mismatches 76; Indels 40; Gaps 9;  
QY 56 GGVVPAALAA--WREALATAEVRPDGFFRRVPHDYHPPESLGFDRCCLLAAMDY 112  
DB 24 GGGVPLAAAPLPFGDLAASGKLRDTP-----NATVPASMDFGNITAAALPAAVLF 74  
QY 113 PGQDSIVIDM-----GTAITIDLLAGGHRGGRILP--GIAMSLRGLHEGTALLPEV 162  
DB 75 PGSPGDVAELLRAAYAAPGRPFTVSRGRGHSTMGQALAAAGVVVHMOSMGGGA--PRI 132  
QY 163 VLNAPAEMLGNDTSNAIQAGVIHLFADALRGAITDFRQYSPQARI----LITGGDAERWQ 218  
DB 133 NVSADGAY-----VDAGGEQLWVDVLRALA--RGVAPRSWTDLHLTVGGTSL--N 180  
QY 219 PGIAGSLYQ 227  
DB 181 AGVSGQTYR 189  
RESULT 6  
VG42\_BPMU  
ID VG42\_BPMU  
AC Q9T1V6;

DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Protein gp42.  
GN 42.  
OS Bacteriophage Mu.  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;  
OC Mu-like viruses.  
OX NCBI\_TaxID=10677;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Morgan G., Hatfull G., Hendrix R.;  
RT "Genome of bacteriophage Mu and comparison with the Haemophilus  
RT influenzae Mu-like prophage Flumu.";  
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: PUTATIVE TAPE MEASURE PROTEIN.  
CC -!- SIMILARITY: TO H.INFLUENZAE HI1514.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF083977; AAF01120.1; -  
SQ SEQUENCE 690 AA; 73595 MW; 21FALACF3F6072E6 CRC64;  
Query Match 7.3%; Score 92; DB 1; Length 690;  
Best Local Similarity 29.1%; Pred. No. 1.7;  
Matches 53; Conservative 12; Mismatches 91; Indels 26; Gaps 7;  
QY 50 REPVALGGVVPAALAAWREALATAEVRPDGFFRRVPHDYHPPESLGFDRCCLLAAA 109  
DB 452 RQPLLLSGPQPLALPAPRPVLA-----PPGVPVPTARPAPLPLPGKSGLLSLAGSAA 506  
QY 110 MDYPGQDSIVIDMGTAITIDLLAGGHFRG-GRILPGIAMSRLRGLHEGTALLPEVVLNAPA 168  
DB 507 QLVGTGTGKLADAGRAV-----GGWFSGIGNKLAGSAIG-RVVTKGAGALGWMKGAGR 559  
QY 169 EM--LGNDTSNAIQAGVIHLFADALRGAITDFRQYSPQARIITGGDAERWQPGIAGSLY 226  
DB 560 ALSRLGGPVMGALQ-----LAPVLMDEQASTHEKAGAGTAGAWLGGAVGSLA 608  
QY 227 QP 228  
DB 609 GP 610  
RESULT 7  
YP65\_MYCTU  
ID YP65\_MYCTU STANDARD; PRT; 583 AA.  
AC Q50733;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein rv2565.  
GN RV2565 OR MT2641 OR MTCY9C4.03C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,



RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence.";  
RL Nature 393:537-544(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CDC 1551 / Oshkosh;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains.";  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO THE UPF0028 (SWS) FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; Z77250; CAB01050.1; -.  
DR EMBL; AE007098; AAK46954.1; -.  
DR TIGR; MT2641; -.  
DR Tuberculist; RV2565; -.  
DR InterPro; IPR002641; Patatin.  
DR InterPro; IPR001423; UPF0028.  
DR InterPro; IPR000595; cNMP\_binding.  
DR Pfam; PF00027; cNMP\_binding; 1.  
DR Pfam; PF01734; Patatin; 1.  
DR SMART; SM00100; cNMP; 1.  
DR PROSITE; PS0042; cNMP\_BINDING\_3; 1.  
DR PROSITE; PS01237; UPF0028; 1.  
KW Hypothetical protein; Complete proteome.  
FT CONFLICT 375 375 G -> D (IN REF. 2).  
SQ SEQUENCE 583 AA; 62123 MW; EC6EE062BAEC62F7 CRC64;  
  
Query Match 7.2%; Score 91.5; DB 1; Length 583;  
Best Local Similarity 24.3%; Pred. No. 1.5;  
Matches 45; Conservative 25; Mismatches 74; Indels 41; Gaps 7;  
  
y 62 ALAAWREALATAEVREPDGFFRRVPHDYHPESIGFDRRCCLLAAMDYPGQDSIVID 121  
b 258 SLATQRPVHLITCLAEPPDPSPWMDRLAPVSHHPANSDGFGALARRIA----- 303  
  
QY 122 MGTAITIDLLAGHFRGGRILPGIAMSRLGLHESFALLPEVVLNAPAEMLGNDTSNAIQA 181  
Db 304 -GRSLGL-VMAGGGARG-----LAHFGVYQELTEAGVVI-----DRFGGTSSGAIAS 348  
  
QY 182 GVIHLFADALRGAITDPRQ-----YSPQARILITGGDAERWQPGIAGSLYQPHLL 231  
Db 349 AAFALGMDA-GDAIAAREFIAGSDPLGDYTIPIALTRGGRVDRLVQGGFTLIEH-L 406  
  
QY 232 LRGFY 236  
Db 407 PRGFF 411  
  
RESULT 8  
DHAL\_PSESP  
ID DHAL\_PSESP STANDARD; PRT; 488 AA.  
AC P33008;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 01-FEB-1996 (Rel. 33, Last annotation update)  
DE Probable aldehyde dehydrogenase (EC 1.2.1.3).  
GN TERPE.  
OS Pseudomonas sp.

OC Bacteria; Proteobacteria.  
OX NCBI\_TaxID=306;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92332528; PubMed=1629218;  
RA Peterson J.A., Lu J.-Y., Geisselsoder J., Graham-Lorence S.,  
RA Carmona C., Witney F., Lorence M.C.;  
RT "Cytochrome P-450terp. Isolation and purification of the protein and  
RT cloning and sequencing of its operon.";  
RL J. Biol. Chem. 267:14193-14203(1992).  
CC -1- FUNCTION: INVOLVED IN A ALPHA-TERPINEOL OXIDATION SYSTEM.  
CC -1- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.  
CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M31440; AAA25995.1; -.  
DR EMBL; S27652; S27652.  
DR PIR; C42971; C42971.  
DR HSSP; P51977; 1BXS.  
DR InterPro; IPR002086; Aldehyde\_dehydr.  
DR Pfam; PF0171; aldedh; 1.  
DR PROSITE; PS00070; ALDEHYDE\_DEHYDR\_CYS; 1.  
DR PROSITE; PS00687; ALDEHYDE\_DEHYDR\_GLU; 1.  
KW Oxidoreductase; NAD.  
FT NP\_BIND 240 245 NAD (ADP PART) (BY SIMILARITY).  
FT ACT\_SITE 262 262 BY SIMILARITY.  
FT ACT\_SITE 296 296 BY SIMILARITY.  
SQ SEQUENCE 488 AA; 52156 MW; 2144AE1EB173C854 CRC64;  
  
Query Match 7.2%; Score 91; DB 1; Length 488;  
Best Local Similarity 28.6%; Pred. No. 1.4;  
Matches 54; Conservative 19; Mismatches 60; Indels 56; Gaps 12;  
  
QY 28 ATSLPPTTEILQOPGLTWLSAPNREPVALGGVWPAALAAWREALATAEVREPDGFFRRV 87  
Db 131 AGSLPLPEVWHEQSATRLSKTVREPL---GVV-VAIVPNYPVSIAFVK-PD-----RRV 180  
  
QY 88 PHDYHPPESLGFDRRCCLLAAMDYPGQDSIVIDMTGTAITIDLLAG-----GH--- 135  
Db 181 GCRKHADSQAGADH-----AADYLRIIGELIRDALPAGVINVLAGNDDLGPWLTGHADV 233  
  
QY 136 ----FRGGRILPGIAMSRLGH---ECTALLPEVVLNAPAEMLGNDTSNAIQAGVIHLFA 188  
Db 234 AKISFTGS-----SVTGRHVARSAAADLKRRLTL-----ELGGND-----AAVV--MA 273  
  
QY 189 DALRGAITD 197  
Db 274 DLSDAVVE 282  
  
RESULT 9  
PAC1\_PSES3  
ID PAC1\_PSES3 STANDARD; PRT; 557 AA.  
AC P15557;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Acylase ACY 1 [Includes: Cephalosporin acylase (EC 3.5.1.-) (GL-7ACA  
DE acylase); Gamma-glutamyltranspeptidase (EC 2.3.2.2) (GGT)].  
GN ACYI.  
OS Pseudomonas sp. (strain SE83).  
OC Bacteria; Proteobacteria.  
OX NCBI\_TaxID=309;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-21 AND 367-387.  
RX MEDLINE=88058804; PubMed=3680178;



```
*RESULT 11
PEX6_HUMAN
ID PEX6_HUMAN STANDARD; PRT; 980 AA.
AC Q13608; Q99476;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Peroxisome assembly factor-2 (PAF-2) (Peroxisomal-type ATPase 1)
DE (Peroxin-6).
GN PEX6 OR PXAAAL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96272151; PubMed=8670792;
RA Yahraus T., Braverman N., Dodt G., Kalish J.E., Morrell J.C.,
RA Moser H.W., Valle D., Gould S.J.;
RA "The peroxisome biogenesis disorder group 4 gene, PXAAAL, encodes a
RT cytoplasmic ATPase required for stability of the PTS1 receptor.";
RL EMBO J. 15:2914-2923(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97094178; PubMed=8940266;
RA Fukuda S., Shimozawa N., Suzuki Y., Zhang Z., Tomatsu S.,
RA Tsukamoto T., Hashiguchi N., Osumi T., Masuno M., Imaizumi K.,
RA Kuroki Y., Fujiki Y., Orii T., Kondo N.;
RA "Human peroxisome assembly factor-2 (PAF-2): a gene responsible for
RT group C peroxisome biogenesis disorder in humans.";
RL Am. J. Hum. Genet. 59:1210-1220(1996).
RN [3]
RP SEQUENCE FROM N.A., AND VARIANTS ZS GLN-812 AND TRP-812.
RX MEDLINE=99335264; PubMed=10408779;
RA Zhang Z., Suzuki Y., Shimozawa N., Fukuda S., Imamura A.,
RA Tsukamoto T., Osumi T., Fujiki Y., Orii T., Wanders R.J.A.,
RA Barth P.G., Moser H.W., Paton B.C., Besley G.T., Kondo N.;
RA "Genomic structure and identification of 11 novel mutations of the
RT PEX6 'peroxisome assembly factor-2' gene in patients with peroxisome
RT biogenesis disorders.";
RL Hum. Mutat. 13:487-496(1999).
CC -1- FUNCTION: INVOLVED IN PEROXISOME BIOSYNTHESIS. REQUIRED FOR
CC STABILITY OF THE PTS1 RECEPTOR.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- DISEASE: DEFECTS IN PEX6 ARE THE CAUSE OF PEROXISOME BIOGENESIS
CC INHERITED DISEASES. THEY ARE CLINICALLY DIVIDED INTO FOUR GROUPS:
CC ZELLWEGER SYNDROME (ZS), NEONATAL ADRENOLEUKODYSTROPHY (NALD),
CC INFANTILE REFSUM DISEASE (IRD), AND CLASSICAL RHIZOMELIC
CC CHONDRODYSPLASIA PUNCTATA (RCDP). ZS IS THE MOST SEVERE PHENOTYPE
CC OF THE PBD AND IS CHARACTERIZED BY THE ABSENCE OF PEROXISOMES AND
CC BY MULTIPLE METABOLIC DEFECTS. PATIENTS WITH ZS HAVE SEVERE
CC NEUROLOGICAL ABNORMALITIES, DYSMORPHIC FACIAL FEATURES,
CC HEPATOMEGALY, AND MULTIPLE RENAL CYSTS; MOST DIE WITHIN 6 MONTHS
CC OF BIRTH. NALD AND IRD ARE SIMILAR TO ZS, BUT PROGRESSIVELY
CC MILDER, WITH SOME IRD PATIENTS SURVIVING BEYOND THE THIRD DECADE
CC OF LIFE. CLASSICAL RCDP, DISTINCT FROM THE ZS SPECTRUM, IS
CC ASSOCIATED WITH A CHARACTERISTIC FACIAL APPEARANCE, RHIZOMELIA,
CC ICHTHYOSIS, AND SEVERE NEUROLOGICAL IMPAIRMENT.
CC -1- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U56602; AAC50655.1; -.
CC EMBL; D83703; BAA12069.1; -.
CC EMBL; AF108098; AAF62564.1; -.
CC EMBL; AF108095; AAF62564.1; JOINED.
```

```
DR EMBL; AF108096; AAF62564.1; JOINED.
DR EMBL; AF108097; AAF62564.1; JOINED.
DR Genew; HGNC:8859; PEX6.
DR MIM; 601498; -.
DR MIM; 214100; -.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003959; AAA_ATPase_cent.
DR InterPro; IPR003960; AAA_sub.
DR Pfam; PF00004; AAA; 2.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00674; AAA; 1.
KW Peroxisome; ATP-binding; Repeat; Disease mutation.
FT NP_BIND 470 477 ATP (POTENTIAL).
FT NP_BIND 744 751 ATP (POTENTIAL).
FT VARIANT 812 812 R -> Q (IN ZS).
FT VARIANT 812 812 /FTID=VAR_007918.
FT VARIANT 812 812 R -> W (IN ZS; ATYPICAL).
FT VARIANT 77 77 /FTID=VAR_007919.
FT CONFLICT 77 77 S -> N (IN REF. 1).
SQ SEQUENCE 980 AA; 104060 MW; 0EC1C2A75CE0038F CRC64;

Query Match 7.0%; Score 88.5; DB 1; Length 980;
Best Local Similarity 21.3%; Pred. No. 5.2;
Matches 54; Conservative 27; Mismatches 89; Indels 83; Gaps 10;

QY 25 TSLPPTTEILQQPGLTWSAPNREPVALGGVV---PAALAAWREALATAEVRPDGFFR 84
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 14 TETPPLAVLLPPGGPW-----PAELGLVLRPAGESPAGALLVAALLEGPDAG--- 63
| | | | | | | | | | | | | | | | | | | | | |
QY 85 RAVPHDYHPPESLGRDRCCLLAAMDYPGQDSIVIDMTAITIDLLAGGHERGRIL-- 142
| | | | | | | | | | | | | | | | | | | | | |
Db 64 -----TEEQG-----PGPPQLLVSRALLRLLALGSGAWVRARVRP 100
| | | | | | | | | | | | | | | | | | | | | |
QY 143 PGIAMSLRGLHEGTALLPEV-----VLNAPAEMLGNDTSNAIQAGVHFLFADALRG 193
| | | | | | | | | | | | | | | | | | | | | |
Db 101 PALGWALLGTSGLPGGLPRVGPPLLVRRGETLPVGPVRV-LETRPALQG---LLPGGTRL 155
| | | | | | | | | | | | | | | | | | | | | |
QY 194 AITDFRQYSPQARILITGGDAERWQP-----GIAGSLY 226
| | | | | | | | | | | | | | | | | | | | | |
Db 156 AVTELR--GRALCPESGDSRRPPPPVSVFVSGTVRRLLQGVLTGDSLGSVRSCL 212
| | | | | | | | | | | | | | | | | | | | | |
QY 227 QPHLLLRGFYLI 239
| | | | | | | | | | | | | | | | | | | | | |
Db 213 RGLGLFQGEWVW 225

RESULT 12
ARGD_MYCTU STANDARD; PRT; 400 AA.
ID ARGD_MYCTU STANDARD; PRT; 400 AA.
AC P94990;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Acetylornithine aminotransferase (EC 2.6.1.11) (ACOAT).
GN ARGD OR RV1655 OR MT1693 OR MTCY06H11.20.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Brown T., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
```



RL Nature 393:537-544(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains.";  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
CC -|- CATALYTIC ACTIVITY: N(2)-acetyl-L-ornithine + 2-oxoglutarate - N-  
CC acetyl-L-glutamate 5-semialdehyde + L-glutamate.  
CC -|- COFACTOR: PYRIDOXAL PHOSPHATE.  
CC -|- PATHWAY: Arginine biosynthesis; fourth step.  
CC -|- SIMILARITY: BELONGS TO CLASS-III OF PYRIDOXAL-PHOSPHATE-DEPENDENT  
CC AMINOTRANSFERASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; Z85982; CAB06649.1; --  
CC EMBL; AE007032; AAK45962.1; --  
CC HSSP; P12995; IQJ3.  
CC TIGR; MT1693; --  
CC TubercuList; Rv1655; --  
CC InterPro; IPR000954; Aminotran\_3.  
CC InterPro; IPR004636; ArgD.  
CC Pfam; PF00202; aminotran\_3; 1.  
CC TIGRFAMS; TIGR00707; argD; 1.  
CC PROSITE; PS00600; AA\_TRANSFER\_CLASS\_3; 1.  
KW transferase; Aminotransferase; Pyridoxal phosphate;  
KW Arginine biosynthesis; Complete proteome.  
FT BINDING 253 253 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
SQ SEQUENCE 400 AA; 40910 MW; 833846D529795019 CRC64;  
  
Query Match 6.8%; Score 86; DB 1; Length 400;  
Best Local Similarity 27.4%; Pred. No. 3;  
Matches 51; Conservative 15; Mismatches 54; Indels 66; Gaps 11;  
  
QY 6 VGNRTLLAHTDGVHFDVSVATSL-----PPTEILQQPGLTWSAPNREP 52  
Db 230 MGRGTGAFFAHQHDGITPDVVTAKGLGGLPIGACLAVGPAEELL-TPGL-----HG 280  
53 VALGGVVPAAALAAWREALATAFEVREPDGFFRRVPHDYHPPESLGGDRRCCLLAAAMDY 112  
Db 281 STEGG-NPVCAAA---ALAVLRVLSAD-GLVRR-----EVLGKSLRHGI--EALGH 325  
QY 113 PGQDSIVIDMGTAITIDLLAGHFRGGRILPGIAMSRLRGLHEG-----TALLPE 161  
Db 326 P-----LID-----HVRGRLLGLGALTAPHAKDAEATARDAGYLVNAAAPD 367  
162 VVLNAP 167  
368 VIRLAP 373  
  
RESULT 13  
BCHD\_RHOCA  
ID BCHD\_RHOCA STANDARD; PRT; 561 AA.  
AC P26175;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Magnesium-chelataase 60 kDa subunit (Mg-protoporphyrin IX chelataase)  
DE (Mg-chelataase subunit D).  
GN BCHD.

OS Rhodobacter capsulatus (Rhodopseudomonas capsulata).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;  
OC Rhodobacter.  
OX NCBI\_TaxID=1061;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SB1003 / St Louis;  
RA Burke D.H., Alberti M., Armstrong G.A., Hearst J.E.;  
RL Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.  
CC -|- FUNCTION: Involved in bacteriochlorophyll biosynthesis; introduces  
CC a magnesium ion into protoporphyrin IX to yield Mg-protoporphyrin  
CC IX.  
CC -|- PATHWAY: Bacteriochlorophyll biosynthesis.  
CC -|- SIMILARITY: BELONGS TO THE MG-CHELATASE SUBUNITS D/I FAMILY.  
CC -|- SIMILARITY: CONTAINS 1 VWFA DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; Z11165; CAA77537.1; --  
CC PIR; S17821; S17821.  
CC InterPro; IPR002035; VWFA\_A.  
CC SMART; SM00327; VWA; 1.  
CC PROSITE; PS50234; VWFA; 1.  
KW Photosynthesis; Bacteriochlorophyll biosynthesis.  
FT DOMAIN 379 559 VWFA.  
FT DOMAIN 237 260 GLU/PRO-RICH.  
FT DOMAIN 239 251 POLY-PRO.  
FT DOMAIN 448 451 POLY-GLY.  
SQ SEQUENCE 561 AA; 59740 MW; A54EBCE1C68FCD2D CRC64;  
  
Query Match 6.8%; Score 86; DB 1; Length 561;  
Best Local Similarity 25.7%; Pred. No. 4.5;  
Matches 53; Conservative 22; Mismatches 75; Indels 56; Gaps 10;  
  
QY 10 RTLLAHTDGVHFDVSVATSLPPT-----ILQQPGLTWSAPNREPVALGGVPAAL 63  
Db 50 RRLPPNVDDGALYGGLDVAETLHSGKPVKGLGGLDRPSVFIPLMAERCTAKLGARLAQAL 109  
QY 64 AAWREAL-ATAEVREPDGFFRRVPHDYHPPESLGGDRRCCLLAAAMDYPGQDSIVIDM 122  
Db 110 DLRQHALLALDEAAEPD-----EALPH-----AVADRLG---LFVDL 143  
QY 123 GTAITIDLLAGHFRGGRILPGIAMSRLRGLHEGTALLPEVVLNAPAEMLGNDTSNAIQAG 182  
Db 144 SEVRSID-----GPGLLPETAQ-----IERARELLPQVQM--PAERVSEIVEGCRQLG 189  
QY 183 VIHLFADALRGAITDFRQYSPQARIL 208  
Db 190 ISSLRAPML--ALT-----AARIL 206  
  
RESULT 14  
TRPE\_AZOBR  
ID TRPE\_AZOBR STANDARD; PRT; 732 AA.  
AC P50872;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Anthranilate synthase (EC 4.1.3.27) [Includes: Glutamine  
DE amidotransferase].  
GN TRPE(G).  
OS Azospirillum brasilense.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;  
OC Azospirillum.  
OX NCBI\_TaxID=192;  
RN [1]  
RP SEQUENCE FROM N.A.







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:46:39 ; Search time 27.2799 Seconds  
(without alignments)  
1177.181 Million cell updates/sec

Title: US-09-813-453A-63  
Perfect score: 1264  
Sequence: 1 MFIAGVNTRTLATHDGV.....AGSLYQPHLLRGFYLRG 241

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

arched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002:\*

1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1264	100.0	241	AAU911179	Pantothenate kinas
2	229.5	18.2	246	AAU911156	Thermotoga maritim
3	227	18.0	257	AAU911174	Pantothenate kinas
4	227	18.0	260	AAU911173	Pantothenate kinas
5	215	17.0	262	AAU911155	Deinococcus radiop
6	203	16.1	258	AAU01243	B. subtilis novel
7	203	16.1	258	AAU911149	Bacillus subtilis
8	190.5	15.1	262	AAU911170	Pantothenate kinas
9	189	15.0	219	AAU911176	Pantothenate kinas
10	185	14.6	259	ABB47661	Listeria monocytog

11	184	14.6	265	23	AAU911151	Streptomyces coeli
12	183.5	14.5	244	23	AAU911168	Pantothenate kinas
13	181	14.3	233	23	AAU911163	Pantothenate kinas
14	180.5	14.3	212	23	AAU911177	Pantothenate kinas
15	179.5	14.2	254	23	AAU911171	Pantothenate kinas
16	178.5	14.1	258	23	AAU911172	Pantothenate kinas
17	174	13.8	257	23	AAU911160	Synechocystis pant
18	172.5	13.6	255	23	AAU911154	Geobacter sulfurre
19	170.5	13.5	273	23	AAU911157	Treponema pallidum
20	169	13.4	250	23	AAU911150	Clostridium acetob
21	168.5	13.3	258	23	AAU911153	Rhodobacter capsul
22	164.5	13.0	229	23	AAU911159	Aquifex aeolicus p
23	159	12.6	272	22	AAG81225	Mycobacterium tube
24	159	12.6	272	23	AAU911152	Bordella pertussis
25	154.5	12.2	267	23	AAU911162	Pantothenate kinas
26	145	11.5	256	23	AAU911175	Pantothenate kinas
27	144.5	11.4	223	19	AAW98422	H. pylori GHPO 344
28	144.5	11.4	223	23	AAU911161	Helicobacter pylor
29	144.5	11.4	223	23	AAU911181	Pantothenate kinas
30	144.5	11.4	248	23	AAU911164	Pantothenate kinas
31	143.5	11.4	242	23	AAU911180	Pantothenate kinas
32	136	10.8	262	23	AAU911158	Borrelia burgdorfe
33	124	9.8	209	23	AAU911165	Pantothenate kinas
34	123.5	9.8	455	20	AAU38617	Neisseria gonorrhoe
35	123.5	9.8	455	21	AAU74908	Neisseria gonorrhoe
36	123.5	9.8	460	23	AAU911167	Pantothenate kinas
37	123.5	9.8	592	20	AAU38618	Neisseria gonorrhoe
38	123.5	9.8	592	21	AAU74911	Neisseria gonorrhoe
39	116	9.2	389	21	AAU74909	Neisseria meningit
40	115.5	9.1	592	20	AAU38615	Neisseria meningit
41	115.5	9.1	592	21	AAU74912	Neisseria meningit
42	115.5	9.1	592	23	AAU911169	Pantothenate kinas
43	112.5	8.9	455	21	AAU74910	Neisseria meningit
44	112.5	8.9	592	20	AAU38616	Neisseria meningit
45	112.5	8.9	592	21	AAU74913	Neisseria meningit

ALIGNMENTS

RESULT 1

AAU911179

ID AAU911179 standard; Protein; 241 AA.

XX AAU911179;

AC AAU911179;

DT 05-JUN-2002 (first entry)

XX Pantothenate kinase (Coax) #17.

DE Pantothenate kinase; Coax; antibiotic; antimicrobial;

XX Pantothenate kinase modulator; coenzyme A; bactericidal compound.

KW Thiobacillus ferrooxidans.

OS WO200215601-A2.

PN 28-FEB-2002.

XX 24-AUG-2001; 2001WO-US26531.

PR 24-AUG-2000; 2000US-227860P.

XX 20-MAR-2001; 2001US-0813453.

PA (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX Yocum RR, Patterson TA;

PI WPI; 2002-269358/31.

XX N-PSDB; ABR54200.

PT Identifying potential antibiotic or antimicrobial agent, comprises contacting composition comprising pantothenate kinase (Coax) protein

PT with test compound and identifying inhibitor of the Coax protein -  
XX  
PS Claim 10; Page 109-110; 128pp; English.  
XX  
CC The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.  
XX  
SQ Sequence 241 AA;

Query Match 100.0%; Score 1264; DB 23; Length 241;  
Best Local Similarity 100.0%; Pred. No. 9e-132;  
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 MIFIAVGNTRTLLAHTHDGVHFDVSVVATSLPPTTEILQQPGLTWLSAPNREPVALGGVVP 60  
1 MIFIAVGNTRTLLAHTHDGVHFDVSVVATSLPPTTEILQQPGLTWLSAPNREPVALGGVVP 60  
61 AALAAWREALATAEVRPDGPFRRVPHDYHPPESLGFDRCCLLAAAMDYPGQDSIVI 120  
61 AALAAWREALATAEVRPDGPFRRVPHDYHPPESLGFDRCCLLAAAMDYPGQDSIVI 120  
121 DMGTAITIDLLAGGHRGGRILPGIAMSRLRGLHEGTALLPEVVLNAPAEMLGNDTSNAIQ 180  
121 DMGTAITIDLLAGGHRGGRILPGIAMSRLRGLHEGTALLPEVVLNAPAEMLGNDTSNAIQ 180  
181 AGVIHLFADALRGAITDFRQYSPQARILITGGDAERWQPGIAGSLYQPHLLRGLFWIR 240  
181 AGVIHLFADALRGAITDFRQYSPQARILITGGDAERWQPGIAGSLYQPHLLRGLFWIR 240  
241 G 241  
241 G 241

RESULT 2  
AAU91156  
ID AAU91156 standard; Protein; 246 AA.  
XX  
AC AAU91156;  
XX  
DT 05-JUN-2002 (first entry)  
Thermotoga maritima pantothenate kinase Coax.  
Pantothenate kinase; Coax; antibiotic; antimicrobial;  
pantothenate kinase modulator; coenzyme A; bactericidal compound.  
Thermotoga maritima.  
WO200216601-A2.  
28-FEB-2002.  
24-AUG-2001; 2001WO-US26531.  
24-AUG-2000; 2000US-227860P.  
20-MAR-2001; 2001US-0813453.  
(OMNI-) OMNIGENE BIOPRODUCTS INC.  
Yocum RR, Patterson TA;  
WPI; 2002-269358/31.  
Identifying potential antibiotic or antimicrobial agent, comprises

PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -  
XX  
PS Claim 10; Page 74-75; 128pp; English.  
XX  
CC The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.  
XX  
SQ Sequence 246 AA;

Query Match 18.2%; Score 229.5; DB 23; Length 246;  
Best Local Similarity 28.2%; Pred. No. 7e-17;  
Matches 71; Conservative 40; Mismatches 110; Indels 31; Gaps 5;  
2 IFIAVGNTRTLLAHTHDGVHFDVSVVATSLPPTTEILQQPGL-TWLSAPNREPVALG--GV 58  
3 LLVDVGNTHSVFSTEDGKTRRWRRLSTGVFQTEDELFSLHPLLDGAMREIKIGIVASV 62  
59 VPAALAA-----WREALATAEVRPDGPFRRVPHDYHPPESLGFDRCC 104  
63 VPTQNTVIERFSQKYFHISPIWVKAK-----NGCVKWNKPNPSEVGADRVAN 109  
105 LLAAAMDYPGQDSIVIDMGTAITIDLLAGGHRGGRILPGIAMSRLRGLHEGTALLPEVVL 164  
110 VVAFVKEY-GKNGIIDMGTAITVDLVVNGSVEGAILPGFFMMVHSLFRGTAKPLVEV 168  
165 NAPAEMLGNDTSNAIQAGVIHLFADALRGAITDFRQYSPQARILITGGDAERWQPGIAGS 224  
169 KPADFVVGKDTENIRLGVVNGSVYALEGIIGRIKEVYGDLPVLTGGQSKIIVKDMIKHE 228  
225 LYQPHLLLRGFY 236  
229 IFDEDLTKGVY 240

RESULT 3  
AAU91174  
ID AAU91174 standard; Protein; 257 AA.  
XX  
AC AAU91174;  
XX  
DT 05-JUN-2002 (first entry)  
Pantothenate kinase (Coax) #12.  
Pantothenate kinase; Coax; antibiotic; antimicrobial;  
pantothenate kinase modulator; coenzyme A; bactericidal compound.  
Chlorobium tepidum.  
WO200216601-A2.  
28-FEB-2002.  
24-AUG-2001; 2001WO-US26531.  
24-AUG-2000; 2000US-227860P.  
20-MAR-2001; 2001US-0813453.  
(OMNI-) OMNIGENE BIOPRODUCTS INC.  
Yocum RR, Patterson TA;  
WPI; 2002-269358/31.  
N-PSDB; ABK54195.



-XX  
PT Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -  
XX  
-PS Claim 10; Page 103-104; 128pp; English.  
XX  
CC The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.  
-XX ? Sequence 257 AA;

Query Match 18.0%; Score 227; DB 23; Length 257;  
Best Local Similarity 32.1%; Pred. No. 1.4e-16;  
Matches 84; Conservative 43; Mismatches 101; Indels 34; Gaps 12;  
QY 2 IFIAVGNTRTLLAHTHDGVHFDV-SVATSLPPTTEILQQPGLTWLSAPNREP--VALGGV 58  
Db 3 LVVDIGNTSTTLA-IFTGDEEPSVESVPSALFADSSSTMREVEFGNMARKHGEPPQAIACSV 61  
QY 59 VPAALA---AWREALATAEV-----REPDGFFRRAVPHDYHPPESLGFDRRCCLLAA 108  
Db 62 VPSATAVGSALLESFVPLTTCCKLRFP---FRL---DYATPHTFGADRLALCAWS 113  
QY 109 AMDYPGQDSIVIDMGTAITIDLL-AGGHERGGRILPGIAMSRLGLHEGTALLPEVVLNAP 167  
Db 114 RHLFSEKPVIAVDIGTAITFDVLDTVGNRGGLIMPIDMMAGALHSRTAQLPQVRIDRP 173  
QY 168 AEMLGNDTSNAIQAGV-----IHLFADALRG-AITDFRQYSPQARILITGGDAERWQP 219  
Db 174 ESLLGRSTTECIKSGVFWGVVKQIGGLVDAIRGDLVRDFGEST--VEVIVTGGNSRIIVP 231  
QY 220 GIAG-SLYQPHLLLRGFLWIR 240  
Db 232 EIGPVSVIDEHAVLRGSDLLR 253

RESULT 4  
AAU91173 standard; Protein; 260 AA.  
AC AAU91173;  
DT 05-JUN-2002 (first entry)  
XX Pantothenate kinase (Coax) #11.  
DE Pantothenate kinase; Coax; antibiotic; antimicrobial;  
XX pantothenate kinase modulator; coenzyme A; bactericidal compound.  
KW Caulobacter crescentus.  
XX WO200216601-A2.  
PN 28-FEB-2002.  
XX 24-AUG-2001; 2001WO-US26531.  
PF 24-AUG-2000; 2000US-227860P.  
XX 20-MAR-2001; 2001US-0813453.  
PR (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX Yocum RR, Patterson TA;  
PI  
XX

DR WPI; 2002-269358/31.  
DR N-PSDB; ABK54194.  
XX  
PT Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -  
XX  
PS Claim 10; Page 102-103; 128pp; English.  
XX  
CC The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.  
XX

SQ Sequence 260 AA;  
Query Match 18.0%; Score 227; DB 23; Length 260;  
Best Local Similarity 29.0%; Pred. No. 1.4e-16;  
Matches 75; Conservative 43; Mismatches 103; Indels 38; Gaps 12;  
QY 1 MIFIAVGNTRTLLAHTHDG----VHFDVSVVATSLPP-----TEILQQPGLTWLSAPNR 50  
Db 2 LLAIEQGNNTMTFA-IHDGASVVAQWRSATESTRTADEYVVLSQLLSMQGLGFRAI--- 57  
QY 51 EPVALGGVVPAALAAWR-----EALATAEVREPDGFFRRAVPHDYHPPESLGFDR 101  
Db 58 DAVIISVVVPSQIFNLNLSRRYFNVEPLVIGENA-----KLGDVRIKPSSEAGADR 110  
QY 102 RCCLLAAAMDYPGQDSIVIDMGTAITIDLLAG-GHERGGRILPGIAMSRLGLHEGTALLP 160  
Db 111 LVNAIGAAMVYPG-PLVVDSGTATTFDIVAADGAFEGGIAPGINLSMQALHEAAAKLP 169  
QY 161 EVVLNAPA--EMLGNDTSNAIQAGVIHLFADALRGAITDFR-QYSPQARILITGGDAERW 217  
Db 170 RIAIQRPAQNRIVGTDVVSAMQSGVFWGYISLIEGLVARIKAEERGEPMTVIATGGVASLF 229  
QY 218 QPGIAGSL--YQPHLLLRG 234  
Db 230 E-GATDSIDHFDSDLTIRG 247

RESULT 5  
AAU91155  
ID AAU91155 standard; Protein; 262 AA.  
XX AAU91155;  
AC AAU91155;  
XX 05-JUN-2002 (first entry)  
DT Deinococcus radiopugnans pantothenate kinase Coax.  
XX Pantothenate kinase; Coax; antibiotic; antimicrobial;  
DE pantothenate kinase modulator; coenzyme A; bactericidal compound.  
KW Deinococcus radiopugnans.  
XX WO200216601-A2.  
PN 28-FEB-2002.  
XX 24-AUG-2001; 2001WO-US26531.  
PF 24-AUG-2000; 2000US-227860P.  
XX 20-MAR-2001; 2001US-0813453.  
PR (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX Yocum RR, Patterson TA;  
PI  
XX









AAU91151 standard; Protein; 265 AA.  
AAU91151;  
05-JUN-2002 (first entry)  
Streptomyces coelicolor pantothenate kinase Coax.  
Pantothenate kinase; Coax; antibiotic; antimicrobial;  
pantothenate kinase modulator; coenzyme A; bactericidal compound.  
Streptomyces coelicolor.  
WO200216601-A2.  
28-FEB-2002.  
24-AUG-2001; 2001WO-US26531.  
24-AUG-2000; 2000US-227860P.  
20-MAR-2001; 2001US-0813453.  
(OMNI-) OMNIGENE BIOPRODUCTS INC.  
Yocum RR, Patterson TA;  
WPI; 2002-269358/31.  
Identifying potential antibiotic or antimicrobial agent, comprises  
contacting composition comprising pantothenate kinase (Coax) protein  
with test compound and identifying inhibitor of the Coax protein -  
Claim 10; Page 69-70; 128pp; English.  
The invention describes assays for identifying a (potential) antibiotic  
comprising contacting an assay composition comprising a pantothenate  
kinase (Coax) protein with a test compound, and determining the ability  
of the test compound to inhibit the activity of the Coax protein, an  
essential enzyme for the production of coenzyme A. Coax protein is a  
valuable target for identifying bactericidal compounds. Coax modulating  
agents can be used in an infectious animal model to determine the  
efficacy, toxicity, or side effects of treatment with such an agent. This  
is the amino acid sequence of a pantothenate kinase (Coax) protein  
described in the invention.  
Sequence 265 AA;  
Query Match 14.6%; Score 184; DB 23; Length 265;  
Best Local Similarity 28.8%; Pred. No. 8.8e-12;  
Matches 76; Conservative 39; Mismatches 113; Indels 36; Gaps 11;  
QY 1 MIFIAVGNFTLLAHTHDG-----VHFDSVSVATSLPPTTEILQ-----QPGLTWLSA 47  
Db 2 LLTIDVGNHTHTVLG-LFDGEDIVEHWRISTDSRRTADEL--AVLQGLMGHPLLGDELG 58  
QY 48 PNREPVALGGVPAALAAWREALATAEVRP---DPGFFRRAVPHDYHPPESLGFDRRC 103  
Db 59 DGIDGIAICATVPSVLHELREVTTRYGDVPAVLVEPG-VKTGVPIITDHPKEVGADRII 117  
QY 104 CLLAAAMDYPQDSIVIDMGTAITIDLL-AGGHFRGGRILPGIAMSRLHGTALLPEV 162  
Db 118 NAAVAVELY-GGPAIVVDFGTATTFDAVSARGEYIGGVIAPGIEISVEALGVKGAQLRKI 176  
QY 163 VLNAPAEMLGNDTSNAIQAGVIHLFADALRGAIT----DFRQYSPQARILITGGDAERWQ 218  
Db 177 EVARPRSVIGKNTVEAMQSGIVYGAGQVDGVVNRMARELADDPDVTVIATGGLA---- 232  
QY 219 PGIAGSL-----YQPHLLLRGFYL 237  
Db 233 PMVLGESSVIDEHEPWLTLMGLRL 256  
RESULT 12

AAU91168  
ID AAU91168 standard; Protein; 244 AA.  
XX  
AC AAU91168;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Pantothenate kinase (Coax) #6.  
XX  
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX  
OS Porphyromonas gingivalis.  
XX  
PN WO200216601-A2.  
XX  
PD 28-FEB-2002.  
XX  
PF 24-AUG-2001; 2001WO-US26531.  
XX  
PR 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX  
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX  
PI Yocum RR, Patterson TA;  
XX  
DR WPI; 2002-269358/31.  
DR N-PSDB; ABK54189.  
XX  
PT Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -  
XX  
PS Claim 8; Page 94-95; 128pp; English.  
XX  
CC The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.  
XX  
SQ Sequence 244 AA;  
Query Match 14.5%; Score 183.5; DB 23; Length 244;  
Best Local Similarity 35.4%; Pred. No. 8.9e-12;  
Matches 51; Conservative 23; Mismatches 65; Indels 5; Gaps 3;  
QY 96 SLGFDRRCCLLAAAMDYPQDSIVIDMGTAITID-LLAGGHFRGGRILPGIAMSRLGLHE 154  
Db 92 TLGADRLAAVVGAAHSLYPNTTELLVIDAGTAITYERSVAEGIYLGNIISPGHLRFAKHL 151  
QY 155 GTALLPEV--VLNAPAEMLGNDTSNAIQAGVIHLFADALRGAITDFRQYSPQARILITG 211  
Db 152 FTGRLPLIDPSGISPKIAEYGSSTEEAITAGVIHGLAGEIDRYIDDLHAKEGRSAVILTG 211  
QY 212 GDAERWQPGI-AGSLYQPHLLLRG 234  
Db 212 GDANYLARIIRSGILIHDPDLVLLG 235  
RESULT 13  
AAU91163  
ID AAU91163 standard; Protein; 233 AA.  
XX  
AC AAU91163;  
XX  
DT 05-JUN-2002 (first entry)



KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX  
OS Bacillus halodurans.  
XX  
PN WO200216601-A2.  
XX  
PD 28-FEB-2002.  
XX  
PF 24-AUG-2001; 2001WO-US26531.  
XX  
PR 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX  
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX  
PI Yocum RR, Patterson TA;  
XX  
DR WPI; 2002-269358/31.  
N-PSDB; ABK54192.  
PT Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein  
XX  
PS Claim 10; Page 100; 128pp; English.  
XX  
CC The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.  
XX  
SQ Sequence 254 AA;  
  
Query Match 14.2%; Score 179.5; DB 23; Length 254;  
Best Local Similarity 26.1%; Pred. No. 2.6e-11;  
Matches 61; Conservative 47; Mismatches 97; Indels 29; Gaps 9;  
  
QY 1 MIFIAVGNTRTLLAHTHDGVHFDVSV-----ATSLPTE-----ILQPGLTWLS 46  
Db : : | | | | | : : | | | | | : : | | | : : | | :  
2 ILVIDVGNNTVTL-----GVYQDETLVHHWRLATSRQKTEDEYAMTVRSLEFDHAGLQF-- 54  
  
47 APNREPVALGGVVPAAALAAWREALATAEVREP---DPGFFRRVAVPHDYHPPESLGFDRRC 103  
Db : : : : | | | : : : : | | : : : : | | : : : : | | :  
55 -QDIDGIVISSVVPMPMFSLQCMCKYFHTVPMIIGPG-IKTGLNICYDNPKEVGAD-RI 111  
  
104 CLLAAAMDYPGQDSIVIDMGTAITIDLL-AGGHFRGGRILPGIAMSLRGLHEGTALLPEV 162  
Db | | : : | : | | | | | : : | | | | | : : | | : : | | :  
112 VNAVAIELYGYPAIVVDFGTATTYCLINEKKQYAGGVIAPGIMISTEALYHRASKLPRI 171  
  
163 VLNAPAEMLGNDTSNAIQAGVIHLFADALRGAITDFR-QYSPQARILITGGDAE 215  
QY : | : : : | : : : | : : : | : : : | : : : | : : : | : : :  
Db 172 EIAKPKQVVGVTNTIDSMQSGIFYGYVSQVDGVVKRMKAQAESEPKVIATGGLAK 225

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:08:46 ; Search time 12.4468 Seconds  
(without alignments)  
2373.336 Million cell updates/sec

Title: US-09-813-453A-10  
Perfect score: 1379  
Sequence: 1 MLLIDVGNHVVFGIQENG.....ARLVPTSLPPATVSGSSGN 273

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues  
Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :				Published Applications_AA.*	
				1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*	
				2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*	
				3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*	
				4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*	
				5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*	
				6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*	
				7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*	
				8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*	
				9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*	
				10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*	
				11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*	
				12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*	
				13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*	
				14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*	

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1379	100.0	273	9 US-09-813-453A-10	Sequence 10, Appl
2	371.5	26.9	255	9 US-09-813-453A-7	Sequence 7, Appli
3	354	25.7	256	9 US-09-813-453A-55	Sequence 55, Appl
4	342.5	24.8	258	9 US-09-813-453A-49	Sequence 49, Appl
5	342.5	24.8	262	9 US-09-813-453A-45	Sequence 45, Appl
6	337.5	24.5	260	9 US-09-813-453A-51	Sequence 51, Appl
7	327.5	23.7	258	9 US-09-813-453A-2	Sequence 2, Appli
8	320.5	23.2	254	9 US-09-813-453A-47	Sequence 47, Appl
9	310	22.5	219	9 US-09-813-453A-57	Sequence 57, Appl
10	293.5	21.3	258	9 US-09-813-453A-6	Sequence 6, Appli
11	291.5	21.1	265	9 US-09-813-453A-4	Sequence 4, Appli
12	288	20.9	233	9 US-09-813-453A-17	Sequence 17, Appl
13	285	20.7	246	9 US-09-813-453A-9	Sequence 9, Appli
14	282.5	20.5	250	9 US-09-813-453A-3	Sequence 3, Appli
15	254	18.4	272	9 US-09-712-363-276	Sequence 276, App
16	254	18.4	272	9 US-09-813-453A-5	Sequence 5, Appli
17	248	18.0	262	9 US-09-813-453A-8	Sequence 8, Appli
18	197.5	14.3	257	9 US-09-813-453A-53	Sequence 53, Appl
19	192.5	14.0	212	9 US-09-813-453A-59	Sequence 59, Appl

20	172.5	12.5	244	9	US-09-813-453A-41	Sequence 41, Appl
21	170.5	12.4	241	9	US-09-813-453A-63	Sequence 63, Appl
22	170	12.3	262	9	US-09-813-453A-11	Sequence 11, Appl
23	147.5	10.7	592	9	US-09-813-453A-22	Sequence 22, Appl
24	147.5	10.7	592	9	US-09-813-453A-43	Sequence 43, Appl
25	146.5	10.6	460	9	US-09-813-453A-39	Sequence 39, Appl
26	136.5	9.9	257	9	US-09-813-453A-13	Sequence 13, Appl
27	136	9.9	248	9	US-09-813-453A-20	Sequence 20, Appl
28	132.5	9.6	249	9	US-09-813-453A-70	Sequence 70, Appl
29	122.5	8.9	229	9	US-09-813-453A-12	Sequence 12, Appl
30	118	8.6	249	9	US-09-813-453A-61	Sequence 61, Appl
31	92.5	6.7	351	9	US-09-991-211-11	Sequence 11, Appl
32	92	6.7	604	10	US-09-758-269-14	Sequence 14, Appl
33	92	6.7	4053	9	US-10-123-155-351	Sequence 351, App
34	88	6.4	242	9	US-09-813-453A-65	Sequence 65, Appl
35	88	6.4	268	10	US-09-815-242-11977	Sequence 11977, A
36	87.5	6.3	3519	9	US-09-808-880-4	Sequence 4, Appli
37	85	6.2	280	9	US-09-712-363-278	Sequence 278, App
38	85	6.2	530	9	US-09-738-626-4188	Sequence 4188, Ap
39	82.5	6.0	390	9	US-09-712-363-192	Sequence 192, App
40	82	5.9	922	9	US-09-855-754-6	Sequence 6, Appli
41	82	5.9	922	9	US-10-227-353-6	Sequence 6, Appli
42	82	5.9	4150	9	US-09-808-880-2	Sequence 2, Appli
43	81.5	5.9	223	9	US-09-895-913A-74	Sequence 74, Appl
44	81.5	5.9	223	9	US-09-813-453A-14	Sequence 14, Appl
45	81.5	5.9	223	9	US-09-813-453A-67	Sequence 67, Appl

ALIGNMENTS

RESULT 1

US-09-813-453A-10  
; Sequence 10, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 273  
; TYPE: PRT  
; ORGANISM: Treponema pallidum  
US-09-813-453A-10

Query Match		100.0%;	Score 1379;	DB 9;	Length 273;
Best Local Similarity		100.0%;	Pred. No. 1.8e-121;		
Matches 273;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MLLIDVGNHVVFGIQENGGRVCVRELFR LAPDARKTQDEYSLLIHALCERAGVGRASL	60		
Db	1	MLLIDVGNHVVFGIQENGGRVCVRELFR LAPDARKTQDEYSLLIHALCERAGVGRASL	60		
QY	61	RDAFISSVVPVLTFTIADAVAQISGVQPVVFGPWPAYEHLVPRIPVRAEIGTDLVANAV	120		
Db	61	RDAFISSVVPVLTFTIADAVAQISGVQPVVFGPWPAYEHLVPRIPVRAEIGTDLVANAV	120		
QY	121	AAVHFRSACVVDCGTALTFTTAVDGTGLIQGVAIAPGLRTAVQSLHTGTQALPLVPLAL	180		
Db	121	AAVHFRSACVVDCGTALTFTTAVDGTGLIQGVAIAPGLRTAVQSLHTGTQALPLVPLAL	180		
QY	181	PDSVLGKDTTHVQAGVVRGTLFVIRAMIAQCQKELGCRCAAVITGGLSRFSSEVDPPP	240		

Db 181 PDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCQKELGCRCAAVITGGLSRLFSSEVDFPP 240  
QY 241 IDAQLTSLGLAHIARLVPTSLPPATVSGSSGN 273  
Db 241 IDAQLTSLGLAHIARLVPTSLPPATVSGSSGN 273

RESULT 2  
US-09-813-453A-7  
; Sequence 7, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 255  
; TYPE: PRT  
; ORGANISM: Geobacter sulfurreducens  
US-09-813-453A-7

Query Match 26.9%; Score 371.5; DB 9; Length 255;  
Best Local Similarity 34.7%; Pred. No. 5e-27;  
Matches 87; Conservative 56; Mismatches 101; Indels 7; Gaps 4;  
QY 1 MLLIDVGNHVVFGIQQENGGRVCVRELFR LAPDARKTQDEYSLLIHALCERAGVGRASL 60  
Db 2 LLVIDVGNNTNIVLGIY---DGERLVRD-WRVSTDKARTTDEYGLINELFRLAGLQDI 57  
QY 61 RDAFISSVVPVLTCTIADAVAQISGVQPVVFGPWAYEHLVRIPEPVRAEIGTDLVANAV 120  
Db 58 RAVIISSVVPPLTGVLERLSGLYFGMRPLVVGPGIKTGMPIQYDNP--REVGA DRIVNAV 115  
QY 121 AAYVHFRSACVVVDCGTALTFTAVDGTGLIQGVAIAPGLRTAVQSLHTGTALPLVPLAL 180  
Db 116 AGYEKYRSLIIVDFGTATTFDYVNRKGEYCGGAIAPGLVISTEALFORASKLPRVDIIR 175  
QY 181 PDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCQKELGCRCAAVITGGLSRLFSSE-VDFP 239  
Db 176 PSIIARNVTNVMQAGIYGYVGLVDELIVTRMKAESKDAPRVIA TGGLASLIAPESKTIE 235  
QY 240 PIDAQLTSLGL 250  
Db 236 AVEEYLTLEGL 246

RESULT 3  
US-09-813-453A-55  
; Sequence 55, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 55  
; LENGTH: 256  
; TYPE: PRT  
; ORGANISM: Clostridium difficile  
US-09-813-453A-55  
Query Match 25.7%; Score 354; DB 9; Length 256;  
Best Local Similarity 34.1%; Pred. No. 2.2e-25;  
Matches 87; Conservative 55; Mismatches 99; Indels 14; Gaps 6;  
QY 1 MLLIDVGNHVVFGI-QGENGGRVCVRELFR LAPDARKTQDEYSLLIHALCERAGVGRAS 59  
Db 2 LLVEDVGNNTNIVLGIYKGDK----LVNYWRIKTRDKTSDEYGLISNLFDDYDNVNISD 56  
QY 60 LRDAFISSVVPVLTCTIADAVAQISGVQPVVFGPWAYEHLVRIPEPVRAEIGTDLVANA 119  
Db 57 IDDIVIISVVVPNVNHSLENFCIKYCKKQPLIVGPGIKTGLNLIKYNP--KQVGADRIVNA 114  
QY 120 VAAVHFRSACVVVDCGTALTFTAVDGTGLIQGVAIAPGLRTAVQSLHTGTALPLVPLA 179  
Db 115 VAGIEKYGAPSLVDFGTATTFTCAISEKGEYLGTTIAPGIIKISSEALFQSAKSLPRVELA 174  
QY 180 LPDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCQKELGCRCAAVI-TGGLSRLFSSE--- 235  
Db 175 KPGMTICKSTVSAMQSGIYGYVGLVDKIISIMKELNCDDVKVIATGGLAKLIASETKS 234  
QY 236 VDFPPIDAQLTSLGL 250  
Db 235 IDY--VDGFLTLEGL 247

RESULT 4  
US-09-813-453A-49  
; Sequence 49, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 49  
; LENGTH: 258  
; TYPE: PRT  
; ORGANISM: Bacillus stearothermophilus  
US-09-813-453A-49

Query Match 24.8%; Score 342.5; DB 9; Length 258;  
Best Local Similarity 31.0%; Pred. No. 2.7e-24;  
Matches 78; Conservative 60; Mismatches 105; Indels 9; Gaps 4;  
QY 1 MLLIDVGNHVVFGI-QGENGGRVCVRELFR LAPDARKTQDEYSLLIHALCERAGVGRAS 59  
Db 2 IFVLVDVGNNTNIVLGVYDGE----LKHWRIETSRKTEDEYGMKALLNHVGLQFSD 56  
QY 60 LRDAFISSVVPVLTCTIADAVAQISGVQPVVFGPWAYEHLVRIPEPVRAEIGTDLVANA 119  
Db 57 IRGIIISVVPPINFALERMCLKYFHKPLIVGPGIKTGLDIKYNP--REVGA DRIVNA 114  
QY 120 VAAVHFRSACVVVDCGTALTFTAVDGTGLIQGVAIAPGLRTAVQSLHTGTALPLVPLA 179  
Db 115 VAGIHLGSLIIVDFGTATTTCYINHEKQYMGGAIAPIGIMISTEALFARA AKLPRIEIA 174  
QY 180 LPDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCQKELGCRCAAVITGGLSRLFSSEVD-F 238







```

Query Match      20.7%; Score 285; DB 9; Length 246;
Best Local Similarity 32.4%; Pred. No. 6.2e-19;
Matches 84; Conservative 43; Mismatches 104; Indels 28; Gaps 9;

QY      2 LLIDVGNHVVFQIENGGRVCVRELFRLAPDARKTQDE-----YSLLIHALCERAGVG 56
      . ||:||||: || | : | : ||: : ||: : ||: : || | : | : |
Db      3 LLVDVGNTSHVSFISITD--GKTERR--WRLSTGVFTQTEDEFLSHLHPLLLGDAMREIKGIG 58

QY      57 RASLRDAFISVVVPLVTKTIADAVAQISGVQPVVFGPWAYEH---LPVRIPEPVRAEIGT 113
      . :||| | | : : : : : : : : : : : : : : : : :
Db      59 -----VASVVPTQNTVIERFSQKYFHISPI----WVKAKNGCVKWNVKNP--SEVGA 104

QY      114 DLVANAVAAVHFRSACVWVDCGTALTFTTAVDGTGLIQGVAIAPGLRTAVQSLHTGTAQL 173
      . ||| || : : : ||| | | | : ||| || | | | ||| : |
Db      105 DRVANVAVFKEYGKNGIIMGTATTVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKL 163

QY      174 PLVPLALPDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCQKELGCRCAAIVTGGLSRFLS 233
      . || : | | : ||| : : ||| | : : : : | : | : | : | : :
Db      164 PLVEVKPADFVVGKDTTEENIRLGVVNGSVVALEGIIGRI-KEYVGDLPVVLGTGQSKIVK 222

QY      234 SEVDFFPIDAQLTSLGLAH 252
      . : | ||| : | : |

```

Db 223 DMIKHEIFDEDLTIKGVYH 241

RESULT 14

US-09-813-453A-3

; Sequence 3, Application US/09813453A

; Patent No. US20020168681A1

; GENERAL INFORMATION:

; APPLICANT: Yocum, R. Rogers

; APPLICANT: Patterson, Thomas A.

; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF

; TITLE OF INVENTION: ANTIBIOTICS

; FILE REFERENCE: OGZ-001

; CURRENT APPLICATION NUMBER: US/09/813,453A

; CURRENT FILING DATE: 2001-03-20

; PRIOR APPLICATION NUMBER: US 60/227,860

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: 09/667,569

; PRIOR FILING DATE: 2000-09-21

; NUMBER OF SEQ ID NOS: 77

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 3

; LENGTH: 250

; TYPE: PRT

; ORGANISM: Clostridium acetobutylicum

US-09-813-453A-3

Query Match 20.5%; Score 282.5; DB 9; Length 250;

Best Local Similarity 29.0%; Pred. No. 1.1e-18;

Matches 69; Conservative 60; Mismatches 98; Indels 11; Gaps 4;

QY 1 MLLIDVGNHSHVFGIQGNGRVCVRELFR LAPDARKTQDEYSLLIHALCERAGVGRASL 60

Db 19 ILVDVGNTNIVLGIYNDTK----LTAEWRLSTDVLRSADEYGIQVMNLFQQDKLDPTLV 74

QY 61 RDAFISSVVPVLT KTIA DAVAQISGVQPVVFGPWAYEHL PVRIPPEPVRAEIGTDLVANAV 120

Db 75 EGVIISSVVPNIMYSLEHMIRKYPKINPLVVGPGIKTGINIKYDNP--KEVGADRIVNAV 132

QY 121 AAYVHFRSACVVD CGTALTFTTAVDGTGLIQGVAIAPGLRTAVQSLHTGTAQLPLVPLAL 180

Db 133 AAHEIYKRSLLIIDFGTATTFCAVRENGDYLGGAI CPGIKVSSEALFEKA AKLPRVELIK 192

QY 181 PDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCQKEL----GRC AAVI-TGGLSRLFS 233

Db 193 PAYAICKNTISSIQSGIVRYRLQVKYLFEKLENLPDGRRTRTSLVLATGGLAKLIN 250

RESULT 15

-09-712-363-276

Sequence 276, Application US/09712363

; Patent No. US20020164588A1

; GENERAL INFORMATION:

; APPLICANT: Eisenberg, David

; APPLICANT: Rotstein, Sergio H.

; APPLICANT: Marcotte, Edward M.

; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND

; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS

; FILE REFERENCE: 07419-032001

; CURRENT APPLICATION NUMBER: US/09/712,363

; CURRENT FILING DATE: 2000-11-13

; PRIOR APPLICATION NUMBER: PCT/US00/02246

; PRIOR FILING DATE: 2000-01-28

; PRIOR APPLICATION NUMBER: 60/179,531

; PRIOR FILING DATE: 2000-02-01

; PRIOR APPLICATION NUMBER: 60/117,844

; PRIOR FILING DATE: 1999-01-29

; PRIOR APPLICATION NUMBER: 60/118,206,

; PRIOR FILING DATE: 1999-02-01

; PRIOR APPLICATION NUMBER: 60/126,593

; PRIOR FILING DATE: 1999-03-26

; PRIOR APPLICATION NUMBER: 60/134,093

; PRIOR FILING DATE: 1999-05-14

; PRIOR APPLICATION NUMBER: 60/134,092

; PRIOR FILING DATE: 1999-05-14

; PRIOR APPLICATION NUMBER: 60/165,124

; PRIOR FILING DATE: 1999-11-12

; PRIOR APPLICATION NUMBER: 60/165,086

; PRIOR FILING DATE: 1999-11-12

; NUMBER OF SEQ ID NOS: 292

SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 276

; LENGTH: 272

; TYPE: PRT

; ORGANISM: Mycobacterium tuberculosis

US-09-712-363-276

Query Match 18.4%; Score 254; DB 9; Length 272;

Best Local Similarity 31.3%; Pred. No. 5.7e-16;

Matches 84; Conservative 42; Mismatches 118; Indels 24; Gaps 7;

QY 1 MLLIDVGNHSHVFGIQGNGRVCVRELFR LAPDARKTQDEYSLLIHALCERAGVGRASL 60

Db 2 LLAI DVRNTHTVVGLLSGMKEHAKVVQWRI RTESEVTADELALTIDGL-----IGEDSE 56

QY 61 R--DAFISSVVPVLT KTIA DAVAQISGVQPVVFGPWAYEHL PVRIPPEPVRAEIGTDLVA 117

Db 57 RLGTAAALSTVPSVLHEVRIMLDQYWPSVPHVLI EPGVRTGIPLLV DNP--KEVGADRIV 114

QY 118 NAVAAYVHFRSACVVD CGTALTFTTAVDGTGLIQGVAIAPGLRTAVQSLHTGTAQLPLVP 177

Db 115 NCLAA YDRFRKAAI VVDFGSSICVDVVS AKGEFLGGAIAPGVQVSSDAAAARSALRRVE 174

QY 178 LALPDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCQKELG-----CRCAAVITGGLSRLF 232

Db 175 LARPRSVVGKNTVECMQAGAVFGFAGLV DGLVGRIRE DVS GFSVDHDAIVATGHTAPLL 234

QY 233 SSE---VDFPPIDAQLTSLGLAH IARLV 257

Db 235 LPELHTVDH--YDQHLTQGL-----RLV 256

Search completed: June 24, 2003, 22:23:58

Job time : 14.4468 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:35:06 ; Search time 22.266 Seconds  
(without alignments)  
2526.317 Million cell updates/sec

Title: US-09-813-453A-10  
Perfect score: 1379  
Sequence: 1 MLLIDVGNHVVFGIQGENG.....ARLVPTSLPPATVSGSSGN 273

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues 671580

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_21.\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

ALIGNMENTS									
RESULT 1									
17	170	12.3	262	16	051477	051477 borrelia bu			
18	151.5	11.0	224	16	Q98Q93	Q98Q93 mycoplasma			
19	147.5	10.7	592	16	Q9JXF1	Q9JXF1 neisseria m			
20	147.5	10.7	592	16	Q9JWI7	Q9JWI7 neisseria m			
21	136.5	9.9	257	16	P74045	P74045 synecocyst			
22	136	9.9	248	16	Q9HWC1	Q9HWC1 pseudomonas			
23	132	9.6	276	16	Q8YQD7	Q8YQD7 anabaena sp			
24	127.5	9.2	295	16	Q8Y2M4	Q8Y2M4 ralstonia s			
25	122.5	8.9	229	16	O67753	O67753 aquifex aeo			
26	105	7.6	380	17	Q9HRS8	Q9HRS8 halobacteri			
27	105	7.6	430	16	Q9A4D7	Q9A4D7 caulobacter			
28	104.5	7.6	5532	2	Q9S0R4	Q9S0R4 streptomyce			
29	104	7.5	1456	2	Q9F636	Q9F636 stigmatella			
30	103.5	7.5	3972	2	Q9S0R8	Q9S0R8 streptomyce			
31	98	7.1	3192	2	Q9L4W4	Q9L4W4 streptomyce			
32	96.5	7.0	569	10	Q9AXZ3	Q9AXZ3 persea amer			
33	95.5	6.9	841	16	Q9ACU8	Q9ACU8 streptomyce			
34	95	6.9	366	2	Q8VP07	Q8VP07 burkholderi			
35	95	6.9	613	2	Q93LN4	Q93LN4 xanthomonas			
36	94	6.8	383	16	O53382	O53382 mycobacteri			
37	93.5	6.8	469	16	Q98K12	Q98K12 rhizobium l			
38	93	6.7	335	16	Q9KEK9	Q9KEK9 bacillus ha			
39	93	6.7	347	16	Q9RV45	Q9RV45 deinococcus			
40	93	6.7	638	2	Q9L8K5	Q9L8K5 mycobacteri			
41	92	6.7	604	10	O24592	O24592 zea mays (m			
42	91.5	6.6	252	10	Q94JD6	Q94JD6 oryza sativ			
43	91.5	6.6	327	16	Q98P49	Q98P49 rhizobium l			
44	91.5	6.6	2512	16	Q10896	Q10896 mycobacteri			
45	91.5	6.6	4077	2	O52820	O52820 amycolatops			

SUMMARIES									
Result No.	Score	Query Match	Length	DB	ID	Description			
1	1379	100.0	273	16	O83446	O83446 treponema p			
2	349.5	25.3	259	16	Q8XHL5	Q8XHL5 clostridium			
3	348.5	25.3	255	16	Q8R7M2	Q8R7M2 thermoanaer			
4	345.5	25.1	259	16	Q92F54	Q92F54 listeria in			
5	344.5	25.0	259	16	Q8YAC5	Q8YAC5 listeria mo			
6	337.5	24.5	261	16	Q9A6Z1	Q9A6Z1 caulobacter			
7	329.5	23.9	258	2	Q9F985	Q9F985 bacillus st			
8	320.5	23.2	254	16	Q9KGH5	Q9KGH5 bacillus ha			
9	316.5	23.0	273	16	Q97EB4	Q97EB4 clostridium			
10	291.5	21.1	265	16	Q9X8N6	Q9X8N6 streptomyce			
11	285	20.7	246	16	Q9WZY5	Q9WZY5 thermotoga			
12	267	19.4	274	16	Q9CD56	Q9CD56 mycobacteri			
13	260.5	18.9	256	16	Q8RFE4	Q8RFE4 fusobacteri			
14	254	18.4	272	16	O06282	O06282 mycobacteri			
15	248	18.0	262	16	Q9RX54	Q9RX54 deinococcus			
16	192.5	14.0	212	2	O32514	O32514 desulfovibr			

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Query Match 100.0%; Score 1379; DB 16; Length 273;  
Best Local Similarity 100.0%; Pred. No. 5.4e-98;  
Matches 273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MLLIDVGNHVVFGIQGENGRVCVRELFRLAPDARKTQDEYSLLIHALCERAGVGRASL 60

















**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:56 ; Search time 7.97518 Seconds  
(without alignments)  
1007.182 Million cell updates/sec

Title: US-09-813-453A-10  
Perfect score: 1379  
Sequence: 1 MLLIDVGNHVVFVGIQGENG.....ARLVPTSLPPATVSGSSGN 273

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92.5	6.7	351	4 US-09-632-947B-11	Sequence 11, Appl
2	91	6.6	829	4 US-09-413-814-105	Sequence 105, App
3	87.5	6.3	3519	4 US-09-428-517-4	Sequence 4, Appli
4	83.5	6.1	1049	3 US-08-772-270A-11	Sequence 11, Appl
5	83.5	6.1	1244	5 PCT-US93-10500-2	Sequence 2, Appli
6	82	5.9	922	4 US-08-460-269C-6	Sequence 6, Appli
7	82	5.9	4150	4 US-09-428-517-2	Sequence 2, Appli
8	81.5	5.9	437	2 US-08-989-925-3	Sequence 3, Appli
9	81.5	5.9	2532	4 US-09-215-694-10	Sequence 10, Appl
10	81	5.9	826	4 US-09-564-805-224	Sequence 224, App
11	81	5.9	826	4 US-09-564-805-226	Sequence 226, App
12	80.5	5.8	437	1 US-08-764-343-1	Sequence 1, Appli
13	80.5	5.8	683	4 US-09-134-001C-5576	Sequence 5576, Ap
14	80.5	5.8	819	4 US-09-651-656-15	Sequence 15, Appl
15	80.5	5.8	819	4 US-09-650-855-15	Sequence 15, Appl
16	80.5	5.8	5087	4 US-09-144-085-1	Sequence 1, Appli
17	80	5.8	434	2 US-08-989-925-1	Sequence 1, Appli
18	79	5.7	712	4 US-09-708-426-9	Sequence 9, Appli
19	78.5	5.7	801	1 US-08-725-012-2	Sequence 2, Appli
20	78.5	5.7	826	4 US-09-564-805-2	Sequence 2, Appli
21	78	5.7	669	4 US-09-071-035-264	Sequence 264, App
22	78	5.7	1638	4 US-09-071-035-258	Sequence 258, App
23	78	5.7	1638	4 US-09-071-035-262	Sequence 262, App
24	78	5.7	1638	4 US-09-071-035-266	Sequence 266, App
25	76.5	5.5	418	4 US-09-030-267-5	Sequence 5, Appli
26	76.5	5.5	3816	4 US-09-428-517-3	Sequence 3, Appli
27	76	5.5	161	2 US-08-403-852D-22	Sequence 22, Appl

28	76	5.5	161	3 US-08-510-646B-23	Sequence 23, Appl
29	76	5.5	161	4 US-09-231-818-22	Sequence 22, Appl
30	75.5	5.5	320	4 US-09-036-987A-14	Sequence 14, Appl
31	75.5	5.5	320	4 US-09-370-700-14	Sequence 14, Appl
32	75.5	5.5	1841	2 US-08-804-227C-6	Sequence 6, Appli
33	75.5	5.5	4545	2 US-08-804-227C-14	Sequence 14, Appl
34	75.5	5.5	4550	2 US-08-804-227C-8	Sequence 8, Appli
35	75.5	5.5	4550	2 US-08-804-198-2	Sequence 2, Appli
36	75	5.4	2595	4 US-09-036-987A-2	Sequence 2, Appli
37	75	5.4	2595	4 US-09-370-700-2	Sequence 2, Appli
38	75	5.4	5215	4 US-09-105-537-2	Sequence 2, Appli
39	75	5.4	5588	4 US-09-036-987A-6	Sequence 6, Appli
40	75	5.4	5588	4 US-09-370-700-6	Sequence 6, Appli
41	74.5	5.4	1562	3 US-09-320-878-3	Sequence 3, Appli
42	74.5	5.4	1562	4 US-09-105-537-35	Sequence 35, Appl
43	74.5	5.4	11877	4 US-09-105-537-6	Sequence 6, Appli
44	74	5.4	565	4 US-08-952-365-8	Sequence 8, Appli
45	73.5	5.3	1580	2 US-08-804-227C-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1  
US-09-632-947B-11  
; Sequence 11, Application US/09632947B  
; Patent No. 6356845  
; GENERAL INFORMATION:  
; APPLICANT: Pharmacia & Upjohn  
; TITLE OF INVENTION: CRYSTALLIZATION AND STRUCTURE DETERMINATION OF  
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS  
; TITLE OF INVENTION: UDP-N-ACETYLENOLPYRUVYLGLUCOSAMINE REDUCTASE (S. aureus  
; TITLE OF INVENTION: MurB)  
; FILE REFERENCE: 268.6241 0101  
; CURRENT APPLICATION NUMBER: US/09/632,947B  
; CURRENT FILING DATE: 2000-08-04  
; PRIOR APPLICATION NUMBER: 60/147,164  
; PRIOR FILING DATE: 1999-08-04  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 11  
; LENGTH: 351  
; TYPE: PRT  
; ORGANISM: Bordetella pertussis  
US-09-632-947B-11

Query Match	6.7%	Score 92.5;	DB 4;	Length 351;
Best Local Similarity	24.0%	Pred. No. 0.041;		
Matches	68;	Conservative	36;	Mismatches 94; Indels 85; Gaps 15;
Qy	31	LAPDARKTQDEYSLLIHALCERAGVGRASLRDAFIS----	SVVPVLTKTITADAVAQISGV	86
Db	11	VAPLAPQAQD-----LRCFNT-LGLASHAPAFVALTEPSQLPALS-ALAPRFRQL----		58
Qy	87	QPVVFGPWAYEHLR-----VRPEPVRAEIGTDLVANAAVYV-----HFR	127	
Db	59	--VVLGGGSNVVLPASIDGLVAQVRLP-----GVRLVQCADAWVVEAAAGENWHGFV	109	
Qy	128	SACVVVDCGTALTFTAVDGTGLIQGVAIAPGLRTAVQSLHTGTAQLPLVPLALPDSVLGK	187	
Db	110	TAC--VDNGW-----DG---LENLALIPGTGGAAPVQNIGAYGVELA-----	146	
Qy	188	DTTHAVQAGVVRGTLFVIRAMIAQCQ-----KELGCRCAAVITGGLSLRFSSEVD	237	
Db	147	DRFHSLTAWDVKGGRWV-EMGAECRFAYRDSFFKHQEPGAWVIGSVRFALPRPQPVL	205	
Qy	238	FPPIDAQLTSLGLAHIAIRLVPTSL-----LPPATVSGSSGN	273	
Db	206	YPDLQRHAALDGAAPTARAVYDAVCAIRRAKLPDPAPVGVNAGS	248	

RESULT 2  
US-09-413-814-105





```

; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1049 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Actinobacillus pleuropneumoniae
;
US-08-772-270A-11
Query Match 6.1%; Score 83.5; DB 3; Length 1049;
Best Local Similarity 20.9%; Pred. No. 2.2;
Matches 53; Conservative 47; Mismatches 112; Indels 41; Gaps 9;
QY 26 RELRLAPDARKTQDEYSLLIHALCERAGVGRASLRDAFI---SSVVPVLTKTITADAVAQ 82
Db 93 RNELEVAHKSGLGTADQFL----GLTERGLTLFAPQLDQFLQKHKSISNVVGSSTGDAVSK 148
Y 83 ISGVQPVVFGPWAYEHLVPRIPEVRAEIGTDLVA-NAVAAYVHERSACVVVDCGTALTF 141
Db 149 LAKSQTIIISG-----IQSVLGTVLGINLNEAIIISGGSELEAEAGVSLAS 194
QY 142 TAVDGTGLIQGVAIAPGLRTAVQSL-----HTGTAQLPLVPLALPDSVLGKDTT 190
Db 195 ELV--SNIAGKTTIDAFITQIQNFGKLAENAKGLGGVGRQLQNISSALSKTGLGLDII 252
QY 191 HAVQAGVVRGTLFVIRAMIAQCQKEL--GCRCAAVITGGLSRLFSSEVDFPPIDAQLTSL 248
Db 253 SLLSGVTRS--FALRNKNASTSTKVAAGFELSNOVIGGITKAVSSYILAQRRLAGLSTT 310
QY 249 GLAHIARLVPTSL 261
Db 311 GPA--AALIASSI 321
```

```

RESULT 5
PCT-US93-10500-2
; Sequence 2, Application PC/TUS93310500
; GENERAL INFORMATION:
; APPLICANT: Chang, Yung-Fu
; TITLE OF INVENTION: Recombinant Vaccine For Procline
; TITLE OF INVENTION: Pleuropneumoniae
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Alan S. Korman
; STREET: 1600 Empire Tower
; CITY: Buffalo
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25.
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/10500
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/972,229
; FILING DATE: 05-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Korman, Alan S.
; REGISTRATION NUMBER: 33,932
; REFERENCE/DOCKET NUMBER: 19603/00001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 716-853-8104
; TELEFAX: 716-853-8109
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1244 amino acids
```

```

; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Actinobacillus pleuropneumoniae
; STRAIN: Serotypes 2, 3, 4, 6 and 8
; INDIVIDUAL ISOLATE: Swine
; CELL TYPE: Gram negative bacterium
;
PCT-US93-10500-2
Query Match 6.1%; Score 83.5; DB 5; Length 1244;
Best Local Similarity 20.9%; Pred. No. 2.9;
Matches 53; Conservative 47; Mismatches 112; Indels 41; Gaps 9;
QY 26 RELRLAPDARKTQDEYSLLIHALCERAGVGRASLRDAFI---SSVVPVLTKTITADAVAQ 82
Db 266 RNELEVAHKSGLGTADQFL----GLTERGLTLFAPQLDQFLQKHKSISNVVGSSTGDAVSK 321
QY 83 ISGVQPVVFGPWAYEHLVPRIPEVRAEIGTDLVA-NAVAAYVHERSACVVVDCGTALTF 141
Db 322 LAKSQTIIISG-----IQSVLGTVLGINLNEAIIISGGSELEAEAGVSLAS 367
QY 142 TAVDGTGLIQGVAIAPGLRTAVQSL-----HTGTAQLPLVPLALPDSVLGKDTT 190
Db 368 ELV--SNIAGKTTIDAFITQIQNFGKLAENAKGLGGVGRQLQNISSALSKTGLGLDII 425
QY 191 HAVQAGVVRGTLFVIRAMIAQCQKEL--GCRCAAVITGGLSRLFSSEVDFPPIDAQLTSL 248
Db 426 SLLSGVTRS--FALRNKNASTSTKVAAGFELSNOVIGGITKAVSSYILAQRRLAGLSTT 483
QY 249 GLAHIARLVPTSL 261
Db 484 GPA--AALIASSI 494
```

```

RESULT 6
US-08-460-269C-6
; Sequence 6, Application US/08460269C
; Patent No. 6197548
; GENERAL INFORMATION:
; APPLICANT: CLARE, JEFFREY J.
; ROMANOS, MICHAEL A.
; TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN YEAST
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Millen, White, Zelano & Branigan, P.C.
; STREET: 2200 Clarendon Blvd., Suite 1400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,269C
; FILING DATE: 02-Jun-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lebovitz, Richard M.
; REGISTRATION NUMBER: 37,067
; REFERENCE/DOCKET NUMBER: Popov-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 243-6333
; TELEFAX: (703) 243-6410
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
```



QY 154 AIAPGLRTAVQSLHTGTAQLPLVPLALPDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCQ 213  
Db 125 ----GAKDSVASTITGV-----MDRTKGAVTGSVEKTKSVVSGSI---N 161  
QY 214 KELGCRCAAVITGGL-SRLFSSEV 236  
Db 162 TVLGSRRMQLVSSGVENALTKSEL 185

RESULT 9  
US-09-215-694-10  
; Sequence 10, Application US/09215694B  
; Patent No. 6391583  
; GENERAL INFORMATION:  
; APPLICANT: Wisconsin Alumni Research Foundation  
; APPLICANT: Hutchinson, Charles R.  
; APPLICANT: Kennedy, Jonathan n.m.i  
; APPLICANT: Park, Cheonseek n.m.i  
; TITLE OF INVENTION: METHOD OF PRODUCING ANTIHYPERCHOLESTEROLEMIC AGENTS  
; FILE REFERENCE: 960296.95718  
; CURRENT APPLICATION NUMBER: US/09/215,694B  
; CURRENT FILING DATE: 1999-12-18  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 2532  
; TYPE: PRT  
; ORGANISM: Aspergillus terreus  
US-09-215-694-10

Query Match 5.9%; Score 81.5; DB 4; Length 2532;  
Best Local Similarity 21.1%; Pred. No. 14;  
Matches 38; Conservative 29; Mismatches 68; Indels 45; Gaps 7;  
QY 6 VGNHVVFGIQGNGRVCVRELRLAPDARKTD-EYSLIHALCERAGVGRASLRDAF 64  
Db 64 VGSMHVRGGHFLDEDPALFDASFFNMSTEVASCMDPQYRLILEVYE-----AL 112  
QY 65 ISSVVPVLTKTIAADAVAQISGVQPVVFGPWAY-----EHLVRIPEP-VRAEIGTDLV 116  
Db 113 EAAGIP-----LEQVSGSKTGVFAGTMYHDYQGSFQRPQPEALPRYFITGNAGT-ML 162  
QY 117 ANAVAAVHFRSACVVDCGTALTFTAVDGTGLIQGVAIAPGLRTAVQSLHTGTAQLPLV 176  
Db 163 ANRVSHFYDLRGPSVSDTACSTTLTA-----LHLAIQSLRAGESDMAIV 207

RESULT 10  
JS-09-564-805-224  
; Sequence 224, Application US/09564805  
; Patent No. 6333403  
; GENERAL INFORMATION:  
; APPLICANT: Tavtigian, Sean V.  
; APPLICANT: Teng, David H.F.  
; APPLICANT: Simard, Jacques  
; APPLICANT: Rommens, Johanna M.  
; APPLICANT: Myriad Genetics, Inc.  
; TITLE OF INVENTION: Chromosome 17p-Linked Prostate Cancer Susceptibility  
; FILE REFERENCE: 2318-258  
; CURRENT APPLICATION NUMBER: US/09/564,805  
; CURRENT FILING DATE: 2000-05-05  
; PRIOR APPLICATION NUMBER: US 60/107,468  
; PRIOR FILING DATE: 1998-11-06  
; PRIOR APPLICATION NUMBER: 09/434,382  
; PRIOR FILING DATE: 1999-11-05  
; NUMBER OF SEQ ID NOS: 240  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 224  
; LENGTH: 826  
; TYPE: PRT  
; ORGANISM: Pan troglodytes  
US-09-564-805-224

Query Match 5.9%; Score 81; DB 4; Length 826;  
Best Local Similarity 23.3%; Pred. No. 3;  
Matches 47; Conservative 19; Mismatches 68; Indels 68; Gaps 8;  
QY 26 RELFRIAP----DARKTQDEYSLLIHALCERAGVGRASLRDAFISSV----- 68  
Db 202 RPLSRLSPERSSDSESNENE-PHLPBGVSQRRGVRDSSLVAFICKLHLKRGNFVLKAK 260  
QY 69 ---VPVLTKTIAADAVAQISGVQPVVFGPWAYEHLVRIPEPVRAEIGTDLVANAVAAYVH 125  
Db 261 EMGLPVGTAIAPIIAAVKDGKSI-----THEGREILAEEL----- 296  
QY 126 FRSACVVVDCGTALTFTAVDGTGLIQGVAIAPGLRTAVQSLHTGTAQLP--LVPLALPDS 183  
Db 297 -----CTPPDPGAAFVVVECPDESFIQ-----PICENATFORYQKADAPVALVVHMAPES 347  
QY 184 VL-----GKDTTHAV 193  
Db 348 VLVDSRYQQWMERFGPDTQHLV 369

RESULT 11  
US-09-564-805-226  
; Sequence 226, Application US/09564805  
; Patent No. 6333403  
; GENERAL INFORMATION:  
; APPLICANT: Tavtigian, Sean V.  
; APPLICANT: Teng, David H.F.  
; APPLICANT: Simard, Jacques  
; APPLICANT: Rommens, Johanna M.  
; APPLICANT: Myriad Genetics, Inc.  
; TITLE OF INVENTION: Chromosome 17p-Linked Prostate Cancer Susceptibility  
; FILE REFERENCE: 2318-258  
; CURRENT APPLICATION NUMBER: US/09/564,805  
; CURRENT FILING DATE: 2000-05-05  
; PRIOR APPLICATION NUMBER: US 60/107,468  
; PRIOR FILING DATE: 1998-11-06  
; PRIOR APPLICATION NUMBER: 09/434,382  
; PRIOR FILING DATE: 1999-11-05  
; NUMBER OF SEQ ID NOS: 240  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 226  
; LENGTH: 826  
; TYPE: PRT  
; ORGANISM: Gorilla gorilla  
US-09-564-805-226

Query Match 5.9%; Score 81; DB 4; Length 826;  
Best Local Similarity 23.3%; Pred. No. 3;  
Matches 47; Conservative 19; Mismatches 68; Indels 68; Gaps 8;  
QY 26 RELFRIAP----DARKTQDEYSLLIHALCERAGVGRASLRDAFISSV----- 68  
Db 202 RPLSRLSPERSSDSESNENE-PHLPBGVSQRRGVRDSSLVAFICKLHLKRGNFVLKAK 260  
QY 69 ---VPVLTKTIAADAVAQISGVQPVVFGPWAYEHLVRIPEPVRAEIGTDLVANAVAAYVH 125  
Db 261 EMGLPVGTAIAPIIAAVKDGKSI-----THEGREILAEEL----- 296  
QY 126 FRSACVVVDCGTALTFTAVDGTGLIQGVAIAPGLRTAVQSLHTGTAQLP--LVPLALPDS 183  
Db 297 -----CTPPDPGAAFVVVECPDESFIQ-----PICENATFORYQKADAPVALVVHMAPES 347  
QY 184 VL-----GKDTTHAV 193  
Db 348 VLVDSRYQQWMERFGPDTQHLV 369

RESULT 12  
US-08-764-343-1  
; Sequence 1, Application US/08764343







**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:33:11 ; Search time 4.74823 Seconds  
(without alignments)  
2384.688 Million cell updates/sec

Title: US-09-813-453A-10  
Perfect score: 1379  
Sequence: 1 MLLIDVGNSHVVGIOGENG.....ARLVPTSLPPATVSGSGN 273

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues  
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	288	20.9	233	1	YACB_BACSU	P37564 bacillus su
2	106.5	7.7	344	1	GUN4_TRIRE	O14405 trichoderma
3	98.5	7.1	679	1	MT01_YEAST	P53070 saccharomyc
4	93.5	6.8	335	1	MREH_BACSU	P39763 bacillus su
5	93.5	6.8	459	1	COBG_PSEDE	P21637 pseudomonas
6	92.5	6.7	351	1	MURB_BORPE	Q9x6y8 bordetella
7	91.5	6.6	497	1	GATA_MYCLE	O33105 mycobacteri
8	90.5	6.6	290	1	MURI_DEIRA	Q9rul0 deinococcus
9	90.5	6.6	327	1	FMT_RALSO	Q8y3a8 ralstonia s
10	90	6.5	747	1	GUND_CELFI	P50400 cellulomona
11	88.5	6.4	836	1	VG26_BPMD2	O64220 mycobacteri
12	88	6.4	384	1	POQE_METEX	P71517 methylobact
13	87.5	6.3	3519	1	OL56_STRAT	Q07017 streptomyc
14	86	6.2	285	1	DDH2_HUMAN	O95865 homo sapien
15	85.5	6.2	489	1	CL28_MYCTU	Q59572 mycobacteri
16	85.5	6.2	545	1	DCIP_AZOBR	P51852 azospirillu
17	85	6.2	280	1	DHP1_MYCTU	O06274 mycobacteri
18	85	6.2	351	1	ID12_RHILO	Q98915 rhizobium l
19	84.5	6.1	397	1	DP38_MYCSM	P52851 mycobacteri
20	84	6.1	270	1	ISPD_STRCO	Q910q8 streptomyc
21	84	6.1	562	1	ATKA_ALIAC	Q9xell alicyclobac
22	83.5	6.1	817	1	MUTS_THECA	Q9zix6 thermus cal
23	83.5	6.1	1008	1	SN14_YEAST	P36048 saccharomyc
24	83.5	6.1	1049	1	RT31_ACTPL	P55130 actinobacil
25	82.5	6.0	232	1	VG65_BPMD2	O64257 mycobacteri
26	82.5	6.0	238	1	HIS4_METVA	P05324 methanococc
27	82.5	6.0	327	1	GLK_ZYMMO	P21908 zymomonas m
28	82.5	6.0	338	1	GPDA_STRPY	P58143 streptococc
29	82.5	6.0	381	1	MRP_MYCTU	O33225 mycobacteri
30	82.5	6.0	468	1	HEM1_MYCTU	Q11139 mycobacteri
31	82.5	6.0	665	1	CYAL_RHIME	P19485 rhizobium m
32	82	5.9	166	1	RS5_BACSU	P21467 bacillus su
33	82	5.9	302	1	COBV_PSEDE	P29936 pseudomonas

34	82	5.9	415	1	DP42_RHILO	Q98jm5 rhizobium l
35	82	5.9	836	1	VG26_BPML5	Q05233 mycobacteri
36	82	5.9	922	1	PERT_BORPA	P24328 bordetella
37	82	5.9	3354	1	CADN_MOUSE	Q99pf4 mus musculu
38	81.5	5.9	277	1	MURI_BRUME	Q8yhk2 brucella me
39	81.5	5.9	437	1	ADFP_HUMAN	Q99541 homo sapien
40	81.5	5.9	505	1	YL21_CAEEL	P34414 caenorhabdi
41	81.5	5.9	519	1	LEUI_XYLFA	Q9pcg3 xylella fas
42	81.5	5.9	633	1	DNAK_AGR5	P50019 agrobacteri
43	81.5	5.9	1205	1	SMC_MYCTU	Q10970 mycobacteri
44	81	5.9	430	1	VG43_BPPF3	P03668 bacterioph
45	81	5.9	530	1	MURE_MYCLE	O69557 mycobacteri

ALIGNMENTS

RESULT 1  
YACB\_BACSU  
ID YACB\_BACSU STANDARD; PRT; 233 AA.  
AC P37564;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein yacB.  
GN YACB.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=96051385; PubMed=7584024;  
RA Ogasawara N., Nakai S., Yoshikawa H.;  
RT "Systematic sequencing of the 180 kilobase region of the Bacillus  
subtilis chromosome containing the replication origin.";  
RL DNA Res. 1:1-14(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,  
Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
Kobayashi Y., Koetter, P., Koningsstein G., Krogh S., Kumano M.,  
Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,  
Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,  
Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
subtilis.";  
RL Nature 390:249-256(1997).  
CC -!- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.





RA Fartmann B., Kramer B., Kramer W.;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO THE GIDA FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; 272758; CAA96953.1; -  
DR SGD; S0003205; MT01.  
DR InterPro; IPR002218; GIDA.  
DR InterPro; IPR004416; GIDA\_sub.  
DR Pfam; PF01134; GIDA; 1.  
DR ProDom; PD003738; GIDA; 1.  
DR TIGRFAMs; TIGR00136; gida; 1.  
DR PROSITE; PS01280; GIDA\_1; 1.  
DR PROSITE; PS01281; GIDA\_2; 1.  
KW Mitochondrion.  
SQ SEQUENCE 679 AA; 75374 MW; 8EA1E47F571BECE9 CRC64;  
  
Query Match 7.1%; Score 98.5; DB 1; Length 679;  
Best Local Similarity 21.6%; Pred. No. 1.1;  
Matches 63; Conservative 37; Mismatches 109; Indels 83; Gaps 12;  
  
QY 22 RVCVRELFLAPDARKTQDEYSLLIHALCERAG-----VGRASLRDA 63  
Db | : | | | | | | | | | | | | | | | | | | | : | : |  
22 RLTISSLTSFQP-TTKTQVVVIGAGHAGEAAASRTGAHTTLTLPSTLDIGKSCNPS 80  
QY 64 FISSVVPVLTKTIDA-----VAQISGVQ-----PVVFGPWA-----YEHLPV 101  
Db | : | | | | | | | | | | | | | | | | | | | : | : |  
81 IGGVGKGLVKEI-DALDGLMGKVTDLAGVQFKMLNRSKGPVAVWGPRQAIDRELYKKYMQ 139  
QY 102 RIPEPVRAEIGTDLVANAVAAYVHFRSACVVVDCGTALTFTAVDGTGLIQGVAIAPGLRT 161  
Db | : | | | | | | | | | | | | | | | | | | | : | : |  
140 RELSDKKAHPNLSLQNKVADLILYDPG-----GHKVIKGVLDGDTQV 184  
QY 162 AVQS--LHTGT---AQLPLVPLALPDSVLGKDTTHAVQAGVVVRGTLFVIRAMIAQCQKEL 216  
Db | : | | | | | | | | | | | | | | | | | | | : | : |  
185 GADQVITTTGTFLSAEIHIGDKRIAAGRIGEPTYGIS-----NTL-----QNEV 229  
QY 217 GRCAAVITGGLSRLFSSEVDFPPIDAQ-----LTLSGLAHARLVPTSL 262  
Db | : | | | | | | | | | | | | | | | | | | | : | : |  
230 GFOLGRUKTGTPTARLAKESIDFSALEVQKGDALPVPMSFLNETVSVEPTKQL 281

RESULT 4  
MREH\_BACSU STANDARD; PRT; 335 AA.  
AC P39763;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DE MREH protein.  
GN MREBH.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168 / JH642;  
RA Kobayashi K., Sato T., Kobayashi Y.;  
RT "The nucleotide sequence analysis of kinC region."  
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=97124187; PubMed=8969500;  
RA Winters P., Caldwell R., Enfield L., Ferrari E.;  
RT "The amps-nprE (124 degrees-127 degrees) region of the Bacillus

subtilis 168 chromosome: sequencing of a 27 kb segment and  
identification of several genes in the area."  
Microbiology 142:3033-3037(1996).  
[3]  
SEQUENCE FROM N.A.  
STRAIN=168;  
Caldwell R.M., Ferrari E.;  
"Sequence analysis of the mobA-amps region of the Bacillus subtilis  
chromosome."  
Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
[4]  
SEQUENCE FROM N.A.  
STRAIN=168;  
MEDLINE=98044033; PubMed=9384377;  
Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,  
Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,  
Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,  
Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
subtilis."  
Nature 390:249-256(1997).  
-!- SIMILARITY: BELONGS TO THE FTSA/MREB FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D37799; BAA07047.1; -  
DR EMBL; AF012285; AAC24922.1; -  
DR EMBL; Z99111; CABI3320.1; -  
DR Subtilist; BGL0987; mreBH.  
DR InterPro; IPR001023; Hsp70.  
DR InterPro; IPR004753; mreB\_Mr1.  
DR ProDom; PD000089; Hsp70; 1.  
DR TIGRFAMs; TIGR00904; mreB; 1.  
KW Complete proteome.  
SQ SEQUENCE 335 AA; 35687 MW; 699784E5531323DD CRC64;  
  
Query Match 6.8%; Score 93.5; DB 1; Length 335;  
Best Local Similarity 26.1%; Pred. No. 1.3;  
Matches 62; Conservative 20; Mismatches 77; Indels 79; Gaps 12;  
  
QY 66.SSVVPVLTKTIAADAVAQISGVQVFGPWAYEHLVPRIPVRAEIGTDL-----VANAV 120  
Db | : | | | | | | | | | | | | | | | | | | | : | : |  
107 SGSTAVERRAISDAVKN-CGAKNV-----HL-----IEEPVAAAGADLPVDEPVAN-- 152



Db 11 VAPLAPQAQD-----LRCFNT-LGLASHAPAFVALTEPSQLPALS-ALAPRFRQL--- 58  
QY 87 QPVVFGPWAYEHL-----VRIPEPVRAEIGTDLVANAVAAYV-----HFR 127  
Db 59 --VVLGGGNNVLPASIDGLVAQVRLP-----GVLVGQCADAWVVEAAAGENWHGFV 109  
QY 128 SACVWVDCGTALTFTAVDGTGLIQGVAIAPGLRTAVQSLHTGTQALPLVPLALPDSVLGK 187  
Db 110 TAC--VDNGW-----DG---LENLALIPGTGAPVQNIGAYGVELA----- 146  
QY 188 DTHAVQAGVVRGTLFVIRAMIAQCO-----KELGCRCAAVITGGLSRLESSEVD 237  
Db 147 DRFHSLTAWDKGGRV-EMGAAECRFAYRDSFFKHQEPGAWVIGSVRFALPRWPQVLD 205  
QY 238 FPPIDAQLTSLGLAHIALRVPTSL-----LPPATVSGSSGN 273  
Db 206 YPDQRHAALDGAAPTARAVYDAVCAIRRAKLDPDPVAVVGNAGS 248

RESULT 7

GATA\_MYCLE STANDARD; PRT; 497 AA.  
ID GATA\_MYCLE STANDARD; PRT; 497 AA.  
AC O33105;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Glutamyl-tRNA(Gln) amidotransferase subunit A (EC 6.3.5.-) (Glu-ADT subunit A).  
DE subunit A).  
GN GATA OR ML1702 OR MLCB637.13.  
OS Mycobacterium leprae.  
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1769;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-TN;  
RX MEDLINE-21128732; PubMed-11234002;  
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,  
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,  
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,  
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
RA Barrell B.G.;  
RT "Massive gene decay in the leprosy bacillus."  
RL Nature 409:1007-1011(2001).  
CC -!- FUNCTION: FURNISHES A MEANS FOR FORMATION OF CORRECTLY CHARGED  
CC GLN-tRNA(GLN) THROUGH THE TRANSAMIDATION OF MISACYLATED GLU-  
CC TRNA(GLN) IN ORGANISMS WHICH LACK GLUTAMINYL-tRNA SYNTHETASE. THE  
CC REACTION TAKES PLACE IN THE PRESENCE OF GLUTAMINE AND ATP THROUGH  
CC AN ACTIVATED GAMMA-PHOSPHO-GLU-tRNA(GLN) (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: ATP + L-glutamyl-tRNA(Gln) + L-glutamine = ADP  
CC + phosphate + L-glutamyl-tRNA(Gln) + L-glutamate.  
CC -!- SUBUNIT: HETEROTRIMER OF A, B AND C SUBUNITS (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE AMIDASE FAMILY.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; Z99263; CAB16428.1; --  
CC DR EMBL; AL583923; CAC30555.1; --  
CC DR Leproma; ML1702; --  
CC DR InterPro; IPR000120; Amidase.  
CC DR InterPro; IPR004412; Gata.  
CC Pfam; PF01425; Amidase; 1.

DR TIGRFAMS; TIGR00132; gata; 1.  
DR PROSITE; PS00571; AMIDASES; 1.  
KW Protein biosynthesis; Ligase; Complete proteome.  
SQ SEQUENCE 497 AA; 51536 MW; D3723D871518BDC7 CRC64;  
Query Match 6.6%; Score 91.5; DB 1; Length 497;  
Best Local Similarity 19.2%; Pred. No. 2.9;  
Matches 66; Conservative 37; Mismatches 108; Indels 133; Gaps 16;  
QY 6 VGNHVVFQIGEN-GGRVCVRELPR-----APDARKTQDEYSLLIHALCER-----A 53  
Db 2 IGLHEIIRLDASTLAAKIVAKELSSVEITQACLDQIEATDDTYRAFLHVAEKALSA 61  
QY 54 GVGRA-----SLRDAFI-----SSV 68  
Db 62 AVDKAVAAGGQLSSTLAGVPLALKDVFTTDMPTTCGSKILQGNHSPYDATVTRLRAAG 121  
QY 69 VPVLTKTADAVAQISGVQPVVFG-----PWAYEHL----- 100  
Db 122 IPILGKTNMDEFAMGSSSTENSAYGPTRNPNWVDRVPGSGGSAALAAFAQPLAIGSDT 181  
QY 101 ---VRIPEPVRAEIGTDLVANAVAAYVHFRSACV-VVDCGTALTFTAVDGTGLIQGVAIA 156  
Db 182 GGSIRQPAALTATVGKPTGYGVSRY--GLVACASSLDQGGPCARTVLD-TALLHAVIAG 238  
QY 157 PGLRTAVQSLHTGTQALPLVPLALPDSVLG-----KDTTHAVQAGVVR-----GT 201  
Db 239 HDARDSTS-----VEAAVDP-IVGAAGAGESGDLHGVRVGVVQRLRGEYQSGV 286  
QY 202 LFVIRAMIAQCQKELGCRCAAVITGGLSRLESSEVDFPPIDAQL 245  
Db 287 LASFQAQAVEQ-----LIALGAT---VSEVDCPHFDYAL 316

RESULT 8

MURI\_DEIRA STANDARD; PRT; 290 AA.  
ID MURI\_DEIRA STANDARD; PRT; 290 AA.  
AC O9RU10;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Glutamate racemase (EC 5.1.1.3).  
GN MURI OR DRI586.  
OS Deinococcus radiodurans.  
OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;  
OC Deinococcaceae; Deinococcus.  
OX NCBI\_TaxID=1299;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-R1;  
RX MEDLINE-20036896; PubMed-10567266;  
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,  
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
RA Fraser C.M.;  
RT "Genome sequence of the radioresistant bacterium Deinococcus  
RT radiodurans R1."  
RL Science 286:1571-1577(1999).  
CC -!- FUNCTION: PROVIDES THE (R)-GLUTAMIC ACID REQUIRED FOR CELL WALL  
CC BIOSYNTHESIS (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: L-glutamate = D-glutamate.  
CC -!- PATHWAY: Peptidoglycan biosynthesis.  
CC -!- SIMILARITY: BELONGS TO THE ASPARTATE/GLUTAMATE RACEMASES FAMILY.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC







```
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L02544; AAA23089.1; -.
DR HSSP; P07986; IEXG.
DR InterPro; IPR001919; Bac_cellose-bind.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003962; FNIII_repeat.
DR InterPro; IPR001547; GH_5.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF00553; CBM_2; 1.
DR PRINTS; PR00014; FNTYPEIII.
DR SMART; SM00060; FN3; 2.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
FT SIGNAL 1 39
FT CHAIN 40 747
FT DOMAIN 448 542 FIBRONECTIN TYPE-III 1.
FT DOMAIN 546 639 FIBRONECTIN TYPE-III 2.
FT ACT_SITE 208 208 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 349 349 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 747 AA; 78936 MW; BD15473C9D8B42BD CRC64;

Query Match
Best Local Similarity 25.4%; Score 90; DB 1; Length 747;
Matches 58; Conservative 21; Mismatches 91; Indels 58; Gaps 11;

QY 17 GNGGRVCVRELRLAPDARKTQDE--YSLLIHALCERAG---VGRASLRDAFISSVVPV 71
Db 388 GDTGG-----LLDDWKTWDEVKYSTMLEPTLKHGKYG-----LDHQVPL 430

QY 72 --LTKTIADAVAQISGVQVVFVGPWAYEHL-----PVRIPVPEVRAEIGTDLVANAVAAYVH 125
Db 431 GVGSTGTSTISQVGGTPTDTAPTAPTGLRAGTPTASTVPLTWSASTDTGGSGVAGYEV 490

QY 126 FRSAQVY-----VDCGTALTFT--AVDGTGLIQGVAIAPGLRTAV----- 163
Db 491 YRGTLVGTGTTATSYVTGTLAADSAYTFVSRVRAKDGAGNTSAASAATARTAGGVDVTAP 550

QY 164 ---QSLHTGTAQLPLVPLALPDSVLGKDTTHAVQAG--VVRGTLFVIR 206
Db 551 SVPTGLTAGTPTATSVPLTWAST---DTGGSGVTGYEVYRGSTLVAR 595

RESULT 11
VG26_BPMD2
ID VG26_BPMD2 STANDARD; PRT; 836 AA.
AC 064220;
AT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE Minor tail protein GP26.
GN Mycobacteriophage D29.
OS Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC unclassified Siphoviridae.
OX NCBI_TaxID=28369;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98300335; Pubmed=9636706;
RA Ford M.E., Sarkis G.J., Belanger A.E., Hendrix R.W., Hatfull G.F.;
RT "Genome structure of mycobacteriophage D29: implications for phage evolution.";
RL J. Mol. Biol. 279:143-164(1998).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U72662; AAB58898.1; -.
DR InterPro; IPR000385; Moaa_NifB_PqqE.
DR Pfam; PF01444; Moaa_NifB_PqqE; 1.
DR PROSITE; PS01305; MOAA_NIFB_PQOE; 1.
KW PQQ; Iron-sulfur.
FT METAL 28 28 IRON-SULFUR (POTENTIAL).
```

```
DR EMBL; AF022214; AAC18467.1; -.
FT INIT_MET 0 0 BY SIMILARITY.
SQ SEQUENCE 836 AA; 86572 MW; 528D12ED0FCB92E6 CRC64;

Query Match
Best Local Similarity 23.0%; Score 88.5; DB 1; Length 836;
Matches 73; Conservative 39; Mismatches 114; Indels 91; Gaps 16;

QY 12 VFGIQGNGGRVCVRELRLAPDARKTQDEYSLLIHALCERAG-----VGRA----- 58
Db 409 VLTWVAETLGTALTALQAIQPMPLPTLVDSFKQLSETLVTSGLPYLPQIGAEAFQIVGAV 468

QY 59 -SLRDAFISSVVPVLTFTIADAVAQI--SGVQPVVFGPWAYEHL-PVRIPEPVRAEIGTD 114
Db 469 IQLAPTIISLLIPAF-QTLIPAIQAQLAPSLVQIV---QAFTKLMPIVIVPV-----VQ 516

QY 115 LVANAVAAYVHFRSACVVVDCGTALTFTAVDGTGLIQGV-----AIAPGL---RTAVQS 165
Db 517 IVINLAAA-----VVQAGASIASFLIGGISRLVGLADCVGAVAEVWGSWSGVQV 567

QY 166 LHTGTAQLPLVPLALPDSVLGKDTTHAVQAG-----VVRGTLFVIRAMIAQC---Q 213
Db 568 VSDVFGQLP-----GKIKSWFDDAGSWLIEAGNNVQGLINGISSMISSAVSKA 616

QY 214 KELGCRCAAVITGGL-----SRLFSSEVDF-----PPIDAQLTSLGLAHIA 254
Db 617 KELASSVKNAVGTGFLGIHSPSRVFAEIGQFTAEFGNGFEFGFQPVIEKAKALA--AELS 674

QY 255 RLVPSTLLPPATVSGSS 271
Db 675 QAMESGVDPSGILAGIS 691
```

```
RESULT 12
PQOE_METEX
ID PQOE_METEX STANDARD; PRT; 384 AA.
AC P71517;
AT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE Coenzyme PQQ synthesis protein E.
GN PQOE.
OS Methylobacterium extorquens.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Methylobacterium group; Methylobacterium.
OX NCBI_TaxID=408;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=AM1 / NCIMB 9133;
MEDLINE=97195805; Pubmed=9043136;
RA Toyama H., Chistoserdova L., Lidstrom M.E.;
RT "Sequence analysis of pqq genes required for biosynthesis of
RT pyrroloquinoline quinone in Methylobacterium extorquens AM1 and the
RT purification of a biosynthetic intermediate.";
RL Microbiology 143:595-602(1997).
CC -!- FUNCTION: REQUIRED FOR COENZYME PYRROLO-QUINOLINE-QUINONE (PQQ)
CC BIOSYNTHESIS.
CC -!- SIMILARITY: BELONGS TO THE MOAA / NIFB / PQOE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U72662; AAB58898.1; -.
DR InterPro; IPR000385; Moaa_NifB_PqqE.
DR Pfam; PF01444; Moaa_NifB_PqqE; 1.
DR PROSITE; PS01305; MOAA_NIFB_PQOE; 1.
KW PQQ; Iron-sulfur.
FT METAL 28 28 IRON-SULFUR (POTENTIAL).
```









GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:26 ; Search time 28.0284 Seconds  
(without alignments)  
1297.879 Million cell updates/sec

Title: US-09-813-453A-10  
Perfect score: 1379  
Sequence: 1 MLLIDVGNHVFGIQENG.....ARLVPTSLPPATVSGSSGN 273

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues  
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002:\*  
1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1379	100.0	273	AAU911157	Treponema pallidum
2	371.5	26.9	255	AAU911154	Geobacter sulfurre
3	354	25.7	256	AAU911175	Pantothenate kinas
4	344.5	25.0	259	ABB47661	Listeria monocytog
5	342.5	24.8	258	AAU911172	Pantothenate kinas
6	342.5	24.8	262	AAU911170	Pantothenate kinas
7	337.5	24.5	260	AAU911173	Pantothenate kinas
8	327.5	23.7	258	AAU01243	B. subtilis novel
9	327.5	23.7	258	AAU911149	Bacillus subtilis
10	320.5	23.2	254	AAU911171	Pantothenate kinas

11	310	22.5	219	23	AAU911176	Pantothenate kinas
12	293.5	21.3	258	23	AAU911153	Rhodobacter capsul
13	291.5	21.1	265	23	AAU911151	Streptomyces coeli
14	288	20.9	233	23	AAU911163	Pantothenate kinas
15	285	20.7	246	23	AAU911156	Thermotoga maritlm
16	282.5	20.5	250	23	AAU911150	Clostridium acetob
17	254	18.4	272	22	AAG81225	Mycobacterium tube
18	254	18.4	272	23	AAU911152	Mycobacterium tube
19	248	18.0	262	23	AAU911155	Deinococcus radiop
20	197.5	14.3	257	23	AAU911174	Pantothenate kinas
21	192.5	14.0	212	23	AAU911177	Pantothenate kinas
22	172.5	12.5	244	23	AAU911168	Pantothenate kinas
23	170.5	12.4	241	23	AAU911179	Pantothenate kinas
24	170	12.3	262	23	AAU911158	Borrelia burgdorfe
25	148	10.7	389	21	AAU74909	Neisseria meningit
26	147.5	10.7	455	21	AAU74910	Neisseria meningit
27	147.5	10.7	592	20	AAU38615	Neisseria meningit
28	147.5	10.7	592	20	AAU38616	Neisseria meningit
29	147.5	10.7	592	21	AAU74912	Neisseria meningit
30	147.5	10.7	592	21	AAU74913	Neisseria meningit
31	147.5	10.7	592	23	AAU911166	Pantothenate kinas
32	147.5	10.7	592	23	AAU911169	Pantothenate kinas
33	146.5	10.6	455	20	AAU38617	Neisseria gonorrh
34	146.5	10.6	455	21	AAU74908	Neisseria gonorrh
35	146.5	10.6	460	23	AAU911167	Pantothenate kinas
36	146.5	10.6	592	20	AAU38618	Neisseria gonorrh
37	146.5	10.6	592	21	AAU74911	Neisseria gonorrh
38	136.5	9.9	257	23	AAU911160	Synechocystis pant
39	136	9.9	248	23	AAU911164	Pantothenate kinas
40	132.5	9.6	249	23	AAU911182	Pantothenate kinas
41	122.5	8.9	229	23	AAU911159	Aquifex aeolicus p
42	118	8.6	249	23	AAU911178	Pantothenate kinas
43	114.5	8.3	189	20	AAU38614	Neisseria meningit
44	104.5	7.6	5532	21	AAB23752	S. avermitilis ave
45	104.5	7.6	5532	22	AAG65267	Streptomyces averm

ALIGNMENTS

RESULT 1  
AAU911157  
ID AAU91157 standard; Protein; 273 AA.  
XX  
AC AAU91157;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Treponema pallidum pantothenate kinase Coax.  
XX  
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX  
OS Treponema pallidum.  
XX  
PN WO200216601-A2.  
XX  
PD 28-FEB-2002.  
XX  
PF 24-AUG-2001; 2001WO-US26531.  
XX  
PR 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX  
(OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX  
PI Yocum HR, Patterson TA;  
XX  
DR WPI; 2002-269358/31.  
XX  
PT Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -

XX Claim 6; Page 75-76; 128pp; English.  
PS The invention describes assays for identifying a (potential) antibiotic  
XX comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
XX described in the invention.

SQ Sequence 273 AA;  
Query Match 100.0%; Score 1379; DB 23; Length 273;  
Best Local Similarity 100.0%; Pred. No. 1.2e-143;  
Matches 273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MLLIDVGNHVVFGIQENGGRVCVRELRLAPDARKTQDEYSLLIHALCERAGVGRASL 60  
Db 1 MLLIDVGNHVVFGIQENGGRVCVRELRLAPDARKTQDEYSLLIHALCERAGVGRASL 60  
QY 61 RDAFISSVVPVLTCTIADAVAQISGVQPVVFGPWAYEHLVRIPEPVRAEIGTDLVANAV 120  
Db 61 RDAFISSVVPVLTCTIADAVAQISGVQPVVFGPWAYEHLVRIPEPVRAEIGTDLVANAV 120  
QY 121 AAYVHFRSACVVVDCGTALTFTAVDGTGLIQGVAIAPGLRTAVQSLHTGTAQLPLVPLAL 180  
Db 121 AAYVHFRSACVVVDCGTALTFTAVDGTGLIQGVAIAPGLRTAVQSLHTGTAQLPLVPLAL 180  
QY 181 PDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCQKELGCRCAAVITGGLSRLESSEVDFPP 240  
Db 181 PDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCQKELGCRCAAVITGGLSRLESSEVDFPP 240  
QY 241 IDAQLTSLGLAHARLVPTSLPPATVSGSSGN 273  
Db 241 IDAQLTSLGLAHARLVPTSLPPATVSGSSGN 273

RESULT 2  
AAU91154  
ID AAU91154 standard; Protein; 255 AA.  
XX AAU91154;  
AC  
XX 05-JUN-2002 (first entry)  
DT Geobacter sulfurreducens pantothenate kinase Coax.  
XX  
W Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX Geobacter sulfurreducens.  
OS  
XX WO200216601-A2.  
PN  
XX 28-FEB-2002.  
PD  
XX 24-AUG-2001; 2001WO-US26531.  
PF  
XX 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX  
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
PA  
XX Yocum RR, Patterson TA;  
PI  
XX WPI; 2002-269358/31.  
DR  
XX Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein

PT with test compound and identifying inhibitor of the Coax protein  
XX Claim 10; Page 72-73; 128pp; English.  
PS The invention describes assays for identifying a (potential) antibiotic  
XX comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
XX described in the invention.

SQ Sequence 255 AA;  
Query Match 26.9%; Score 371.5; DB 23; Length 255;  
Best Local Similarity 34.7%; Pred. No. 2.1e-32;  
Matches 87; Conservative 56; Mismatches 101; Indels 7; Gaps 4;  
QY 1 MLLIDVGNHVVFGIQENGGRVCVRELRLAPDARKTQDEYSLLIHALCERAGVGRASL 60  
Db 2 LLVIDVGNNTNVLGIY---DGERLVDR-WRVSTDKARTTDEYGILINELFRAGLGLDQI 57  
QY 61 RDAFISSVVPVLTCTIADAVAQISGVQPVVFGPWAYEHLVRIPEPVRAEIGTDLVANAV 120  
Db 58 RAVIISVVVPVLTCTIADAVAQISGVQPVVFGPWAYEHLVRIPEPVRAEIGTDLVANAV 115  
QY 121 AAYVHFRSACVVVDCGTALTFTAVDGTGLIQGVAIAPGLRTAVQSLHTGTAQLPLVPLAL 180  
Db 116 AGYEKYRTSLIIVDFGTATTFDYVNRKGEYCGGAIAPGLVISTEALFORASKLPRVDIIR 175  
QY 181 PDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCQKELGCRCAAVITGGLSRLESSE-VDFP 239  
Db 176 PSATIIARNTVNSMQAGIYYGYVGLVDEIVTRMKAESKDAPRVATGGLASLIAPESKTIE 235  
QY 240 PIDAQLTSLGL 250  
Db 236 AVEEYLTLEGL 246

RESULT 3  
AAU91175  
ID AAU91175 standard; Protein; 256 AA.  
XX AAU91175;  
AC  
XX 05-JUN-2002 (first entry)  
DT Pantothenate kinase (Coax) #13.  
XX  
DE Pantothenate kinase; Coax; antibiotic; antimicrobial;  
XX pantothenate kinase modulator; coenzyme A; bactericidal compound.  
KW Clostridium difficile.  
XX  
XX WO200216601-A2.  
PN  
XX 28-FEB-2002.  
PD  
XX 24-AUG-2001; 2001WO-US26531.  
PF  
XX 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX  
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
PA  
XX Yocum RR, Patterson TA;  
PI  
XX WPI; 2002-269358/31.  
DR N-PSDB; ABK54196.  
XX







KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX Caulobacter crescentus.  
OS WO200216601-A2.  
PN 28-FEB-2002.  
XX 24-AUG-2001; 2001WO-US26531.  
PD 24-AUG-2000; 2000US-227860P.  
PF 20-MAR-2001; 2001US-0813453.  
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
PA Yocum RR, Patterson TA;  
XX WPI; 2002-269358/31.  
XX N-PSDB; ABK54194.  
XX Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -  
XX  
PS Claim 10; Page 102-103; 128pp; English.  
XX The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.  
XX  
SQ Sequence 260 AA;  
Query Match 24.5%; Score 337.5; DB 23; Length 260;  
Best Local Similarity 32.9%; Pred. No. 1.2e-28;  
Matches 85; Conservative 53; Mismatches 111; Indels 9; Gaps 5;  
QY 1 MLLIDVGNHVVFGIQGNGGRCVRELFRLAPDARKTQDEYSLLIHALCERAGVGRASL 60  
Db 2 LLAIEQGNTNTMFAL---HDGASWVAQ-WRSATESTRTADHEYVWVLSQLLSMQGLGFRAI 57  
JY 61 RDAFISSVVPVLTCTIADAVAQISGVQPVVFGPWAYEHLVRIPEPVRAEIGTDLVANAV 120  
Db 58 DAVIISVVVQSFNRLNLSRRYFNVEPLVIGENAKLIDVRIEKP--SEAGADRLVNAI 115  
QY 121 AAYVHERSACVVVDCGTALTFTTAVDGTGLIQGVAIAPGLRTAVQSLHTGTALPLVPLAL 180  
Db 116 GAAMVYPGVLVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMQALHEAAKLPRIAQR 175  
QY 181 P--DSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCQKELGCRCAAVITGGLSRLFSSEVD- 237  
Db 176 PAGNRIVGTDTVSAMQSGVFWGYISLIEGLVARIKAEERGEPMVTVIATGGVASLFEATDS 235  
QY 238 FPPIDAQLTSLGLAHAR 255  
Db 236 IDHFDSDLTIRGLLEIYR 253  
RESULT 8  
AAU01243  
ID AAU01243 standard; Protein; 258 AA.  
XX  
AC AAU01243;  
XX  
DT 18-JUL-2001 (first entry)  
XX

DE B. subtilis novel pantothenate kinase encoded by the gene coax.  
XX Pantothenate kinase; coax; pantothenate biosynthesis; vitamin B5;  
KW nutritional supplement; panto-compound; pantoate.  
KW  
XX Bacillus subtilis.  
OS  
XX WO200121772-A2.  
PN  
XX 29-MAR-2001.  
PD  
XX 21-SEP-2000; 2000WO-US255993.  
PF  
XX 21-SEP-1999; 99US-0400494.  
PR 07-JUN-2000; 2000US-0210072.  
PR 28-JUL-2000; 2000US-0221836.  
PR 24-AUG-2000; 2000US-0227860.  
XX (OMNI-) OMNIGENE BIOPRODUCTS.  
PA Yocum RR; Patterson TA, Hermann T, Pero JG;  
XX  
PI WPI; 2001-218644/22.  
XX N-PSDE; AAS00984.  
DR  
XX New recombinant microorganism which overexpress a Bacillus subtilis  
PT pantothenate biosynthetic enzyme, useful for the high yield production  
PT of panto-compounds such as pantothenate and pantoate -  
XX  
PS Example 14; Fig 23; 292pp; English.  
XX The sequence represents a novel B. subtilis pantothenate kinase (encoded  
CC by gene coax), an enzyme of the pantothenate biosynthetic pathway.  
CC Pantothenate, also known as vitamin B5, is used as a nutritional  
CC supplement in mammals and humans. The invention concerns methods of  
CC producing recombinant microorganisms overexpressing at least one Bacillus  
CC subtilis pantothenate biosynthetic enzyme. The microorganisms and methods  
CC of producing them are useful for producing a panto-compound such as  
CC pantothenate or pantoate, which is a nutritional requirement for  
CC livestock and humans. The methods are also useful for the identification  
CC of pantothenate kinase modulators. Panto-compounds are produced at a  
CC significantly higher yield than prior art methods and can be produced  
CC independent of the need to feed precursors which decreases expense.  
XX  
SQ Sequence 258 AA;  
Query Match 23.7%; Score 327.5; DB 22; Length 258;  
Best Local Similarity 29.9%; Pred. No. 1.5e-27;  
Matches 75; Conservative 64; Mismatches 105; Indels 7; Gaps 4;  
QY 1 MLLIDVGNHVVFGIQGNGGRCVRELFRLAPDARKTQDEYSLLIHALCERAGVGRASL 60  
Db 2 LLVIDVGNNTNTVLGVY--HDGK--LEYHWRIETSRHKTEDEFGMILRSFLDHSGLMFEQI 57  
QY 61 RDAFISSVVPVLTCTIADAVAQISGVQPVVFGPWAYEHLVRIPEPVRAEIGTDLVANAV 120  
Db 58 DGIISVVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNP--KEVGADRIVNAY 115  
QY 121 AAYVHERSACVVVDCGTALTFTTAVDGTGLIQGVAIAPGLRTAVQSLHTGTALPLVPLAL 180  
Db 116 AAHLYGNPLIVVDFGTATTTCYCIDENKQYMGGAIAPIGITISTEALYSRAAKLPRIETR 175  
QY 181 PDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCQKELGCRCAAVITGGLSRLFSSEVD-FP 239  
Db 175 PDNIIGNTVSAMQSGILFGYGVQGVGIVKRMKQAKQDLKVIATGGLAPLIANESDCID 235  
QY 240 PIDAQLTSLGL 250  
Db 236 IVDPFLLTKGL 246  
RESULT 9  
AAU91149

```

ID  AAU91149 standard; Protein; 258 AA.
XX
AC  AAU91149;
XX
DT  05-JUN-2002 (first entry)
XX
DE  Bacillus subtilis pantothenate kinase Coax.
XX
KW  Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW  pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS  Bacillus subtilis.
XX
PN  WO200216601-A2.
XX
PD  28-FEB-2002.
XX
PF  24-AUG-2001; 2001WO-US26531.
XX
PR  24-AUG-2000; 2000US-227860P.
R  20-MAR-2001; 2001US-0813453.
X
PA  (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI  Yocum RR, Patterson TA;
XX
DR  WPI; 2002-269358/31.
DR  N-PSDB; ABK54168.
XX
PT  Identifying potential antibiotic or antimicrobial agent, comprises
PT  contacting composition comprising pantothenate kinase (Coax) protein
PT  with test compound and identifying inhibitor of the Coax protein -
XX
PS  Claim 10; Page 67-68; 128pp; English.
XX
CC  The invention describes assays for identifying a (potential) antibiotic
CC  comprising contacting an assay composition comprising a pantothenate
CC  kinase (Coax) protein with a test compound, and determining the ability
CC  of the test compound to inhibit the activity of the Coax protein, an
CC  essential enzyme for the production of coenzyme A. Coax protein is a
CC  valuable target for identifying bactericidal compounds. Coax modulating
CC  agents can be used in an infectious animal model to determine the
CC  efficacy, toxicity, or side effects of treatment with such an agent. This
CC  is the amino acid sequence of a pantothenate kinase (Coax) protein
CC  described in the invention.
XX
SQ  Sequence 258 AA;

Query Match 23.7%; Score 327.5; DB 23; Length 258;
Best Local Similarity 29.9%; Pred. No. 1.5e-27;
Matches 75; Conservative 64; Mismatches 105; Indels 7; Gaps 4;

QY 1 MLLIDVGNHSHVVGIOGNGRVCVRELFR LAPDARKTQDEYSLLIHALCERAGVGRASL 60
Db 2 LLVIDVGNNTVLGVY--HDGK--LEYHWRITSRHKTEDEFGMILRSLFDHSGLMFEQI 57
QY 61 RDAFISSVVPVLT KTIA DAVAQISGVQPVVFGPWAYEHL PVRIPPEVRAEIGTDLVANAV 120
Db 58 DGIISVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLN IKYDNP--KEYGADRIVNAV 115
QY 121 AAYVHFRSACVVVDCGTALTFTAVDGTGLIQGVAIAPGLRTAVQSLHTGTAQLPLVPLAL 180
Db 116 AAHLVGNPLIVDFGTATTTCYIDENKQYMGGAIA PGITISTEALYSRAKLPRIEITR 175
QY 181 PDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCQKELGCRCAAVITGGLSRLESSEVD-FP 239
Db 176 PDNIIGKNTVSAMQSGILFGYVGQVEGIVKRMKQAKQDLKVIATGGGLAPLANESDCID 235
QY 240 PIDAQLTSLGL 250
Db 236 IVDPELT LKGL 246
```

```

RESULT 10
AAU91171
ID  AAU91171 standard; Protein; 254 AA.
XX
AC  AAU91171;
XX
DT  05-JUN-2002 (first entry)
XX
DE  Pantothenate kinase (Coax) #9.
XX
KW  Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW  pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS  Bacillus halodurans.
XX
PN  WO200216601-A2.
XX
PD  28-FEB-2002.
XX
PF  24-AUG-2001; 2001WO-US26531.
XX
PR  24-AUG-2000; 2000US-227860P.
PR  20-MAR-2001; 2001US-0813453.
XX
PA  (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI  Yocum RR, Patterson TA;
XX
DR  WPI; 2002-269358/31.
DR  N-PSDB; ABK54192.
XX
PT  Identifying potential antibiotic or antimicrobial agent, comprises
PT  contacting composition comprising pantothenate kinase (Coax) protein
PT  with test compound and identifying inhibitor of the Coax protein -
XX
PS  Claim 10; Page 100; 128pp; English.
XX
CC  The invention describes assays for identifying a (potential) antibiotic
CC  comprising contacting an assay composition comprising a pantothenate
CC  kinase (Coax) protein with a test compound, and determining the ability
CC  of the test compound to inhibit the activity of the Coax protein, an
CC  essential enzyme for the production of coenzyme A. Coax protein is a
CC  valuable target for identifying bactericidal compounds. Coax modulating
CC  agents can be used in an infectious animal model to determine the
CC  efficacy, toxicity, or side effects of treatment with such an agent. This
CC  is the amino acid sequence of a pantothenate kinase (Coax) protein
CC  described in the invention.
XX
SQ  Sequence 254 AA;

Query Match 23.2%; Score 320.5; DB 23; Length 254;
Best Local Similarity 31.4%; Pred. No. 8.9e-27;
Matches 80; Conservative 57; Mismatches 103; Indels 15; Gaps 5;

QY 1 MLLIDVGNHSHVVGIOGNGRVCVRELFR LAPDARKTQDEYSLLIHALCERAGVGRASL 60
Db 2 LLVIDVGNNTVLGVYQDE---TLVHHWRLATSRQKTEDEYAMTVRSLFDHAGLQFQDI 57
QY 61 RDAFISSVVPVLT KTIA DAVAQISGVQPVVFGPWAYEHL PVRIPPEVRAEIGTDLVANAV 120
Db 58 DGIIVSSVPPMFMFSLQEMCKKYFHVTPMIIGPIKTLGN IKYDNP--KEVGADRIVNAV 115
QY 121 AAYVHFRSACVVVDCGTALTFTAVDGTGLIQGVAIAPGLRTAVQSLHTGTAQLPLVPLAL 180
Db 116 AAIELYGPVAVVDFGTATTTCYLINEKKQYAGGVIA PGIMISTEALYHRASKLPRIEIAK 175
QY 181 PDSVLGKDTTHAVQAGVVRGTLF---VIRAMIAQCQKELGCRCAAVITGGLSRLESSEV 236
Db 176 PKQVVGNTIDSMQSGIFGYVSQVDGVVKRMKAQAESE----PKVIATGGLAKLIGTES 231
QY 237 D-FPPIDAQLTSLGL 250
Db 232 ETIDVIDSFLT LKGL 246
```



RESULT 13  
AAU91151  
ID AAU91151 standard; Protein; 265 AA.  
XX AC AAU91151;  
XX DT 05-JUN-2002 (first entry)  
XX DE Streptomyces coelicolor pantothenate kinase Coax.  
XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
XX KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX OS Streptomyces coelicolor.  
XX PN WO200216601-A2.  
XX PD 28-FEB-2002.  
XX F 24-AUG-2001; 2001WO-US26531.  
X 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX PI Yocum RR, Patterson TA;  
XX DR WPI; 2002-269358/31.  
XX PT Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -  
XX Claim 10; Page 69-70; 128pp; English.  
XX The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.  
XX Sequence 265 AA;

Query Match 21.1%; Score 291.5; DB 23; Length 265;  
Best Local Similarity 33.1%; Pred. No. 1.5e-23;  
Matches 89; Conservative 47; Mismatches 108; Indels 25; Gaps 9;  
QY 1 MLLIDVGNHSHVFGI-QGNGGRVVCVRELFRAPDARKTQDEYSLLIHALCERAGVGRAS 59  
Db 2 LLTIDVGNTHFVLGLFDGED-----IVEHWRISTDSSRTADELAVLLQGLMGHPLLGD 56  
QY 60 LRDAF----ISSVVPVLTCTIADAVAQISG-VQPVVFGPWAYEHLPPVRIPEPVRAEIGTD 114  
Db 57 LGDGDGIAICATVPVSVLHELREVTTRYGDPVPAVLVEPGVKTGVPILTDHP--KEVGAD 114  
QY 115 LVANAVAAVHFRSACVWVDCGTALTFTAVDGTGLIQGVAIAPGLRTAVQSLHTGTALP 174  
Db 115 RIINAVAAVELYGGPAIVWDFGTATTDFDARSARGEYIGGVIAPGIEISVEALGVKAQLR 174  
QY 175 LVPLALPDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCQKELG---CRCAAVITGGLSRL 231  
Db 175 KIEVARPRSVIGKNTVEAMQSGIVYGAGQVDGVVNRMARLADDDPDDVTVIATGGLAPM 234  
QY 232 F---SSEVDFPIDAQLTSLGLAHIARLV 257  
Db 235 VLGESSVID--EHEPWLTLMLG-----RLV 257

RESULT 14  
AAU91163  
ID AAU91163 standard; Protein; 233 AA.  
XX AC AAU91163;  
XX DT 05-JUN-2002 (first entry)  
XX DE Pantothenate kinase (Coax) #1.  
XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
XX KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX OS Bacillus subtilis.  
XX PN WO200216601-A2.  
XX PD 28-FEB-2002.  
XX PF 24-AUG-2001; 2001WO-US26531.  
XX PR 24-AUG-2000; 2000US-227860P.  
XX PR 20-MAR-2001; 2001US-0813453.  
XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX PI Yocum RR, Patterson TA;  
XX DR WPI; 2002-269358/31.  
XX DR N-PSDB; ABK54169.  
XX PT Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -  
XX Disclosure; Page 81-82; 128pp; English.  
XX The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.  
XX Sequence 233 AA;  
Query Match 20.9%; Score 288; DB 23; Length 233;  
Best Local Similarity 29.7%; Pred. No. 3.1e-23;  
Matches 62; Conservative 56; Mismatches 85; Indels 6; Gaps 3;  
QY 1 MLLIDVGNHSHVFGI-QGNGGRVVCVRELFRAPDARKTQDEYSLLIHALCERAGVGRASL 60  
Db 2 LLVIDVGNNTVLGVY--HDGK--LEYHWRIETSRHKTEDEFGMILRSFLDHSGLMFEQI 57  
QY 61 RDAFISSVVPVLTCTIADAVAQISG-VQPVVFGPWAYEHLPPVRIPEPVRAEIGTDLVANAV 120  
Db 58 DGIISVVPVPPIMFALERMCTKYFHIEPIQVPGMKTGLNKKYDNP--KEVGADRIVNAV 115  
QY 121 AAYVHFRSACVWVDCGTALTFTAVDGTGLIQGVAIAPGLRTAVQSLHTGTALPLVPLAL 180  
Db 116 AAHLIYGNPLIVDFGTATTTCYIDENKQYMGGAIPGITISTEALYSRAAKLPRIETR 175  
QY 181 PDSVLGKDTTHAVQAGVVRGTLFVIRAMI 209  
Db 176 PDNIIGKNTVSAMQSGILFYGVEGIV 204





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.  
OM protein - protein search, using sw model  
Run on: June 24, 2003, 21:36:04 ; Search time 9.95745 Seconds  
(without alignments)  
2635.685 Million cell updates/sec

Title: US-09-813-453a-10  
Perfect score: 1379  
Sequence: 1 MLLIDVGNSHVFGIQGENG.....ARLVPTSLPPATVSGSSGN 273

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1379	100.0	273	2 D71326	conserved hypothet
2	345.5	25.1	259	2 AF1464	conserved hypothet
3	344.5	25.0	259	2 AF1102	conserved hypothet
4	337.5	24.5	261	2 B87489	transcription acti
5	320.5	23.2	254	2 F83660	hypothetical prote
6	316.5	23.0	273	2 E97293	probable transcrip
7	291.5	21.1	265	2 T36391	hypothetical prote
8	288	20.9	233	2 S66100	conserved hypothet
9	285	20.7	246	2 D72320	conserved hypothet
10	267	19.4	274	2 H86937	conserved hypothet
11	254	18.4	272	2 A70955	hypothetical prote
12	248	18.0	262	2 E75516	conserved hypothet
13	170	12.3	262	2 F70165	conserved hypothet
14	151.5	11.0	224	2 A99571	conserved hypothet
15	147.5	10.7	592	2 B81009	Bira protein/Bvg a
16	147.5	10.7	592	2 H82031	probable biotin-[a
17	136.5	9.9	257	2 S75559	hypothetical prote
18	136	9.9	248	2 H83111	hypothetical prote
19	132	9.6	276	2 AI2292	hypothetical prote
20	122.5	8.9	229	2 E70465	hypothetical prote
21	105	7.6	380	2 D84214	NADH dehydrogenase
22	105	7.6	430	2 A87608	hypothetical prote
23	98.5	7.1	679	2 S64258	hypothetical prote
24	94	6.8	383	2 C70845	probable nagA prot
25	93.5	6.8	335	2 C69660	cell-shape determi
26	93.5	6.8	459	2 B36145	cobG protein - Pse
27	93	6.7	335	2 C83755	cell-shape determi
28	93	6.7	347	2 H75427	S-layer-like array
29	92	6.7	604	2 T04351	viviparous-14 prot

30	91.5	6.6	497	2 T45406	probable amidase (
31	91.5	6.6	2512	2 E70751	probable nrp prote
32	91.5	6.6	4077	2 T17484	hypothetical prote
33	90.5	6.6	334	2 T35939	probable transport
34	90.5	6.6	369	2 C75378	glutamate racemase
35	90.5	6.6	487	2 D87546	benzaldehyde dehyd
36	90	6.5	517	2 B87644	4-coumarate-CoA li
37	90	6.5	747	2 B47093	cellulase (EC 3.2.
38	39.5	6.5	309	2 G83044	lipase LipC PA4813
39	38.5	6.4	302	2 G70614	hypothetical prote
40	38.5	6.4	837	2 H72802	minor tail subunit
41	38.5	6.4	3413	2 T17467	rifamycin polyketi
42	88	6.4	242	2 A82637	conserved hypothet
43	88	6.4	268	2 A83130	conserved hypothet
44	88	6.4	359	2 AB3407	phosphoribosylform
45	87.5	6.3	3519	2 S43048	polyketide synthas

ALIGNMENTS

RESULT 1  
D71326

conserved hypothetical protein TP0431 - syphilis spirochete  
C;Species: Treponema pallidum subsp. pallidum (syphilis spirochete)  
C;Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 20-Jun-2000  
C;Accession: D71326

R;Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.;  
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.;  
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.

Science 281, 375-388, 1998

A;Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.  
A;Reference number: A71250; MUID:98332770; PMID:9665876

A;Accession: D71326

A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA

A;Residues: 1-273 <COL>

A;Cross-references: GB:AE001220; GB:AE000520; NID:g3322705; PIDN:AAC65417.1; PID:g3.

A;Experimental source: strain Nichols

C;Genetics:

A;Gene: TP0431

C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 100.0%; Score 1379; DB 2; Length 273;  
Best Local Similarity 100.0%; Pred. No. 2.7e-108;  
Matches 273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MLLIDVGNSHVFGIQGNGRVCVRELFR LAPDARKTQDEYSLIHALCERAGVGRASL	60
Db	1	MLLIDVGNSHVFGIQGNGRVCVRELFR LAPDARKTQDEYSLIHALCERAGVGRASL	60
QY	61	RDAFISSVVPVLT KTIA DAVAQISGVQP VVFGPWAYEHL PVRIPERAEIGTDLVANAV	120
Db	61	RDAFISSVVPVLT KTIA DAVAQISGVQP VVFGPWAYEHL PVRIPERAEIGTDLVANAV	120
QY	121	AAVHFERSACVVDCGTALTFTTAVDGTGLIQGVAIAPGLRTAVQSLHTGTAQLPLVPLAL	180
Db	121	AAVHFERSACVVDCGTALTFTTAVDGTGLIQGVAIAPGLRTAVQSLHTGTAQLPLVPLAL	180
QY	181	PDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCQKELGCRCAAVITGGLSRLFSSEVDPPP	240
Db	181	PDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCQKELGCRCAAVITGGLSRLFSSEVDPPP	240
QY	241	IDAQLTSLGLAHIA RLVP TSLPPATVSGSSGN	273
Db	241	IDAQLTSLGLAHIA RLVP TSLPPATVSGSSGN	273

RESULT 2  
AF1464

conserved hypothetical protein lin0253 [imported] - Listeria innocua (strain Clip11:  
C;Species: Listeria innocua  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001

C:Accession: AF1464  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
A:Title: Comparative genomics of *Listeria* species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AF1464  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-259 <GLA>  
A:Cross-references: GB:AL592022; PIDN:CAC95486.1; PID:gl6412682; GSPDB:GN00178  
A:Experimental source: strain Clip11262  
C:Genetics:  
A:Gene: lin0253  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 25.1%; Score 345.5; DB 2; Length 259;  
Best Local Similarity 32.6%; Pred. No. 1.3e-21;  
Matches 84; Conservative 53; Mismatches 110; Indels 11; Gaps 4;

Qy 1 MLLIDVGNHSHVFGIQENGGRVCVRELFRLAPDARKTQDEYSLLIHALCERAGVGRASL 60  
Db 2 ILVIDVGNCTCTGVYKEQK---LLRHWMTTDRHRTSDELGMTVNLFFSYANLTPSDI 57

Qy 61 RDAFISSVVPVLTCTIADAVAQISGVQPVVFGPWAYEHLVPRIPVRAEIGTDLVANAV 120  
Db 58 QGIIISSVPPIMHAMETMCVRYFNIRPLIVGPGIKTGLNLKVDNP--REIGSDRIVNAV 115

Qy 121 AAYVHFRSACVVDCGTALTFTTAVDGTGLIQGVAIAPGLRTAVQSLHTGTAQLPLVPLAL 180  
Db 116 AASEEYGTPIVVDFGTATTCYIDVAGVYQGGAIAPIGIMISTEALYNRAAKLPRVDIAE 175

Qy 181 PDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCQKELGCRCAAVITGGLSRLF---SSEVD 237  
Db 176 SSQIIKSTVASMQAGIFYGFGQCEGIIAEMKKQSNTPSPVVVATGGLARMITEKSSAVD 235

Qy 238 FPPIDAQLTSLGLAHIAH 255  
Db 236 I--LDPFLTGLGELLYR 251

Query Match 24.5%; Score 337.5; DB 2; Length 261;  
Best Local Similarity 32.9%; Pred. No. 6.1e-21;  
Matches 85; Conservative 53; Mismatches 111; Indels 9; Gaps 5;

Qy 1 MLLIDVGNHSHVFGIQENGGRVCVRELFRLAPDARKTQDEYSLLIHALCERAGVGRASL 60  
Db 3 LLAIEQNGTNTMFAT---HDGASWVAQ-WRSATESTRTADEYVWVLSQLSMQGLGFRAI 58

Qy 61 RDAFISSVVPVLTCTIADAVAQISGVQPVVFGPWAYEHLVPRIPVRAEIGTDLVANAV 120  
Db 59 DAVIISVVVPQSIFNLRLNRRYFNVEPLVIGENAKLIGIDVRIEKP--SEAGADRLVNAI 116

Qy 121 AAYVHFRSACVVDCGTALTFTTAVDGTGLIQGVAIAPGLRTAVQSLHTGTAQLPLVPLAL 180  
Db 117 GAAMVYGPPLVVIDSGTATTFDIIAADGAFEGGIIAPGINLSMQALHEAAKLPRIAIQR 176

Qy 181 P--DSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCQKELGCRCAAVITGGLSRLFSSSEVD- 237  
Db 177 PAGNRIVGTDTVSAMQSGVFWGYISLIEGLVARIKAEERGEPTVIATGGVASLFEATDS 236

Qy 238 FPPIDAQLTSLGLAHIAH 255  
Db 237 IDHFDSDLTIRGLLEIYR 254

RESULT 3  
AF1102  
conserved hypothetical protein lmo0221 [imported] - *Listeria* monocytogenes (strain EGD-e  
C:Species: *Listeria* monocytogenes  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
Accession: AF1102  
Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
A:Title: Comparative genomics of *Listeria* species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AF1102  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-259 <GLA>  
A:Cross-references: GB:NC\_003210; PIDN:CAD00748.1; PID:gl6409586; GSPDB:GN00177  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo0221  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 25.0%; Score 344.5; DB 2; Length 259;  
Best Local Similarity 32.2%; Pred. No. 1.6e-21;  
Matches 83; Conservative 56; Mismatches 108; Indels 11; Gaps 4;

Qy 1 MLLIDVGNHSHVFGIQENGGRVCVRELFRLAPDARKTQDEYSLLIHALCERAGVGRASL 60

Db 2 ILVIDVGNCTCTGVYKEQK---LLKHWMTTDRHRTSDELGMTVNLFFSYANLTPSDI 57

Qy 61 RDAFISSVVPVLTCTIADAVAQISGVQPVVFGPWAYEHLVPRIPVRAEIGTDLVANAV 120

Db 58 QGIIISSVPPIMHAMETMCVRYFNIRPLIVGPGIKTGLNLKVDNP--REIGSDRIVNAV 115

Qy 121 AAYVHFRSACVVDCGTALTFTTAVDGTGLIQGVAIAPGLRTAVQSLHTGTAQLPLVPLAL 180

Db 116 AASEEYGTPIVVDFGTATTCYIDVAGVYQGGAIAPIGIMISTEALYNRAAKLPRVDIAE 175

Qy 181 PDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCQKELGCRCAAVITGGLSRLF---SSEVD 237

Db 176 SSQIIKSTVSSMQAGIFYGFGQCEGIIAEMKKQSNASPVVVATGGLARMITEKSSAVD 235

Qy 238 FPPIDAQLTSLGLAHIAH 255

Db 236 I--LDPFLTGLGELLYR 251

RESULT 4  
B87489  
transcription activator, probable Baf family [imported] - *Caulobacter crescentus*  
C:Species: *Caulobacter crescentus*  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
C:Accession: B87489  
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; K  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser,  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of *Caulobacter crescentus*.  
A:Reference number: AB7249; MUID:21173698; PMID:11259647  
A:Accession: B87489  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-261 <STO>  
A:Cross-references: GB:AE005673; NID:gl3423392; PIDN:AAK23910.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CC1935  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 24.5%; Score 337.5; DB 2; Length 261;  
Best Local Similarity 32.9%; Pred. No. 6.1e-21;  
Matches 85; Conservative 53; Mismatches 111; Indels 9; Gaps 5;

Qy 1 MLLIDVGNHSHVFGIQENGGRVCVRELFRLAPDARKTQDEYSLLIHALCERAGVGRASL 60

Db 3 LLAIEQNGTNTMFAT---HDGASWVAQ-WRSATESTRTADEYVWVLSQLSMQGLGFRAI 58

Qy 61 RDAFISSVVPVLTCTIADAVAQISGVQPVVFGPWAYEHLVPRIPVRAEIGTDLVANAV 120

Db 59 DAVIISVVVPQSIFNLRLNRRYFNVEPLVIGENAKLIGIDVRIEKP--SEAGADRLVNAI 116

Qy 121 AAYVHFRSACVVDCGTALTFTTAVDGTGLIQGVAIAPGLRTAVQSLHTGTAQLPLVPLAL 180

Db 117 GAAMVYGPPLVVIDSGTATTFDIIAADGAFEGGIIAPGINLSMQALHEAAKLPRIAIQR 176

Qy 181 P--DSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCQKELGCRCAAVITGGLSRLFSSSEVD- 237

Db 177 PAGNRIVGTDTVSAMQSGVFWGYISLIEGLVARIKAEERGEPTVIATGGVASLFEATDS 236

Qy 238 FPPIDAQLTSLGLAHIAH 255

Db 237 IDHFDSDLTIRGLLEIYR 254

RESULT 5  
F83660  
hypothetical protein BH0086 [imported] - *Bacillus halodurans* (strain C-125)  
C:Species: *Bacillus halodurans*  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C:Accession: F83660  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.;







R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998

A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A: Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome

A: Reference number: A70500; MUID: 98295987; PMID: 9634230

A: Accession: A70955

A: Status: preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-272 <COL>

A: Cross-references: GB: 295557; GB: AL123456; NID: g3242276; PIDN: CAB08944.1; PID: g2113976

A: Experimental source: strain H37Rv

C: Genetics:

A: Gene: Rv3600c

C: Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match	18.4%;	Score 254;	DB 2;	Length 272;
Best Local Similarity	31.3%;	Pred. NO. 6.5e-14;		
Matches	84;	Conservative 42;	Mismatches 118;	Indels 24; Gaps 7;
QY	1	MLLIDVGNSHVFGIQENGGRVCVRELFRLPADARKTQDEYSLLIHALCERAGVGRASL	60	
Db	2	LLAIDVRNTHTVGLLSGMKEHAKVVQWRIRTESEVTADELALTIDGL-----IGEDSE	56	
QY	61	R---DAFISSVVPVLTKTIADAVAQISGVQPVVFGFWAYEHLPVRIPEPVRAEIGTDLVA	117	
Db	57	RLTGTAALSTVPSVLHEVRIMLDQYWPSPVPHVLIIEGVRTGIPLLVDNP--KEVGADRV	114	
QY	118	NAVAAYVHERSACVVVDCGTALTFTAVDGTGLIQGVAIAPGLRTAVQSLHTGTQALPLVP	177	
Db	115	NCLAAAYDRFRAAIVDFGSSICVDVVSAKGEFLGGAIAFGVQVSSDAAAARSAAALRRVE	174	
QY	178	LALPDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCQKELG-----CRCAAVITGGLSLRF.	232	
Db	175	LARPRSVVGKNTVECMQAGAVFGAGLVGDLVGRIRREDVSGFSVDHDAIVATGHTAPLL	234	
QY	233	SSE---VDFPPIDAQLTSLGLAHIARLV	257	
Db	235	LPELHTVVDH--VDQHLTQGL-----RLV	256	

RESULT 12  
E75516  
conserved hypothetical protein - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 20-Jun-2000  
C:Accession: E75516  
R.White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, R.J.;  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896; PMID:10567266  
A:Accession: E75516  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-262 <WHI>  
A:Cross-references: GB:AE001905; GB:AE000513; NID:g6458144; PIDN:AAF10040.1; PID:g6458144  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR0461  
A:Map position: 1  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

```

QY 59 SLRDAFISSVVPVLTKTADAVAQISQVQVVFQWAVEHL P-VRIPEPVRAEIGTDLVA 117
      | : | | | : : | : | : : | : | | : : | : | :
Db 63 ----AVLSSVAPPVGENYALAKRHFMDAFV---SAENLPDVTVELDTPGSVGADRLC 115
      | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 118 NAVAA--YVHFSACVVVDCGTALTFTTAVDGTGLIQGVAIAPGLRTAVQSLHTGTQALPL 175
      | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 116 NLFGAEKYLGGLDYAVVVDFTSTNFDVVGRRRFLGGILATGAQVSADALFARAACLPR 175
      | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 176 VPLALPDSVLGKDTTHAVQAGVVRGTLFVIRAMTAQCQKELGCRCAAVITGGLSRL---F 232
      | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 176 ITLQAPETAIGKNTVHALQSLGVFGYAEWVDGLLRIRRAELPGEAAVAVATGGFSRTVQGI 235
      | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 233 SSEVDFPPIDAQLTSLGLAHI 253
      | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 236 CQEIDY--YDETTLRLGLVEL 254
      | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 13
F70165
conserved hypothetical protein BB0527 - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C:Accession: F70165
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.;
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.;
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch,
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943; PMID:9403685
A:Accession: F70165
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-262 <KLE>
A:Cross-references: GB:AE001154; GB:AE000783; NID:g2688431; PIDN:AAC66882.1; PID:g2
A:Experimental source: strain B31

```

Query Match	12.3%;	Score 170;	DB 2;	Length 262;
Best Local Similarity	25.3%;	Pred. No. 7e-07;		
Matches	64;	Conservative 50;	Mismatches 123;	Indels 16; Gaps 6;
QY	2	LLIDVGNShVVFGIQENGGRVCVRELFR LAPDARKTQDE-YSLLIHALCERAGVGRASL	60	
Db	9	LIIDIGNTSIAFAFKDNQ----VNLFIKMTNMLRDEVYSFFEENF-----DFNV	57	
QY	61	RDAFISSVVPVLTKTADAVAQISGVQPVVFGPWAYEHLFVRIPEPVRAEIGTDLVANAV	120	
Db	58	NKVFISVVPIILNETFKNVIFSEFFKIKPLFIGFDLNYDLTFNPYKSDKFLLGSDVFANLV	117	
QY	121	AAYVHFR-SACVVVDCGTALTFTAVDGTGLIQGVAIAPGLRTAVQSLHTGTAQLPLVPLA	179	
Db	118	AAIENYSFENVLVDLGTACTIFAVSRQDGILGGIINSGLPLNFNSLLDNAYLIKFFPIS	177	
QY	180	LPDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCQKELGCRCAAVITGGLSRLFSS--EVD	237	
Db	178	TPNNLLERTTSGSVNSGLFYQYKYLIEGVYRDIKQMYKKKNLIITGGNADLILSLIEIE	237	
QY	238	FPPIDAQLTSLGL	250	
Db	238	F-IFNIHLTVEGV	249	

RESULT 14  
A99571  
conserved hypothetical protein. MYPV\_4730 [imported] - Mycoplasma pulmonis (strain  
C;Species: Mycoplasma pulmonis  
C;Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 03-Aug-2001  
C;Accession: A99571  
R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer,  
Nucleic Acids Res. 29, 2145-2153, 2001  
A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma  
A;Reference number: A99512; MUID:21267165; PMID:11353084  
A;Accession: A99571









Db 68 IGMSPGVVDRDKGTIGAYNLNWKTLQPIKQIEKALGIPFFIDNDANVAALGERWMGA 127  
QY 105 -DRVANVAVFVKEYGKNGIIDMGTA-----TTVDL----- 134  
Db 128 GDNQPDVVFMTLGTGVGGGIVAEGKLLHGVAGAAGELGHITVDFDQPISTCGKKGCLT 187  
QY 135 -----VVN-----GSYEGGAILPGFFMMVHSLFRGTAKLPLVEVKPAD---FVVGKDT 179  
Db 188 VASATGIVNLTRRYADEYEGDAALK---RLIDNGEEVTAKTVFDLAKEGDDLALIVYRNF 244  
QY 180 EENIRLG-----VVNGSVYAL-EGIIIGRIKEVYGD--LPVVLTTGGQSKIIVKD 223  
Db 245 SRYLGIACANIGSILNPSTIVIGGVSAAAGEFLQGVQKVYDENSFPQVRTSTKLALVLI 304  
QY 224 MIKHEIF 230  
Db 305 CLNHSFF 311

RESULT 4  
US-09-110-910A-4  
; Sequence 4, Application US/09110910A  
; Patent No. 6025175  
; GENERAL INFORMATION:  
; APPLICANT: Burnham, Martin K.  
; TITLE OF INVENTION: NOVEL GLUCOSE KINASE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dechert Price & Rhoads  
; STREET: 997 Lenox Drive, Building 3, Suite 210  
; CITY: Lawrenceville  
; STATE: NJ  
; COUNTRY: USA  
; ZIP: 08543  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/110,910A  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/896,083  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bloom, Allen  
; REGISTRATION NUMBER: 29,135  
; REFERENCE/DOCKET NUMBER:  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 609-520-3214  
; TELEFAX: 609-520-3259  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 333 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-110-910A-4

Query Match 7.5%; Score 96; DB 3; Length 333;  
Best Local Similarity 21.5%; Pred. No. 0.014;  
Matches 66; Conservative 42; Mismatches 115; Indels 84; Gaps 14;  
QY 5 VDVGNTHSVFSI-TEDGKTFRRWRLSTGVFQ-----TED--ELFSLHPLLLGDAMREIKG 56  
Db 8 IDLGTSIKFAILTAGEIQKWSIKTNILDEGSHVDDMTESIQHRLDLLGLAAADFQ 67  
QY 57 IGVAS---VWPTQNTVIERFSQKYFHISPIWVKAKNGC-VKWNKPNSEVGA----- 104

Db 68 IGMSPGVVDRDKGTIGAYNLNWKTLQPIKQIEKALGIPFFIDNDANVAALGERWMGA 127  
QY 105 -DRVANVAVFVKEYGKNGIIDMGTA-----TTVDL----- 134  
Db 128 GDNQPDVVFMTLGTGVGGGIVAEGKLLHGVAGAAGELGHITVDFDQPISTCGKKGCLT 187  
QY 135 -----VVN-----GSYEGGAILPGFFMMVHSLFRGTAKLPLVEVKPAD---FVVGKDT 179  
Db 188 VASATGIVNLTRRYADEYEGDAALK---RLIDNGEEVTAKTVFDLAKEGDDLALIVYRNF 244  
QY 180 EENIRLG-----VVNGSVYAL-EGIIIGRIKEVYGD--LPVVLTTGGQSKIIVKD 223  
Db 245 SRYLGIACANIGSILNPSTIVIGGVSAAAGEFLQGVQKVYDENSFPQVRTSTKLALVLI 304  
QY 224 MIKHEIF 230  
Db 305 CLNHSFF 311

RESULT 5  
US-09-710-099-2  
; Sequence 2, Application US/09710099  
; Patent No. 6441154  
; GENERAL INFORMATION:  
; APPLICANT: Donoho, Gregory  
; APPLICANT: Hilbun, Erin  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Nehls, Michael  
; APPLICANT: Friedrich, Glenn  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sands, Arthur T.  
; TITLE OF INVENTION: No. 6441154el Human Proteases and  
; FILE REFERENCE: Polynucleotides Encoding the Same  
; CURRENT APPLICATION NUMBER: US/09/710,099  
; CURRENT FILING DATE: 2000-11-10  
; PRIOR APPLICATION NUMBER: US 60/165,260  
; PRIOR FILING DATE: 1999-11-12  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO: 2  
; LENGTH: 351  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-09-710-099-2

Query Match 7.1%; Score 90.5; DB 4; Length 351;  
Best Local Similarity 22.7%; Pred. No. 0.063;  
Matches 40; Conservative 30; Mismatches 81; Indels 25; Gaps 6;  
QY 24 RRWRLSTGVF-----QTEDELFSLHPLLLGDAMREIKGIGVASVVPQTNTVIERFSQKYF 78  
Db 125 RRLERSTNSFSYSSYHTLEEIYSWIDNFVMEHSDIVSKIQIGNSFENQSIILVKFSTGGS 184  
QY 79 HISPIWVKAKNGCVKWNKPNSEVGADRVANVAVFVKEYGKNGIIDMGTAFT--VDLVV 136  
Db 135 RHPAIWIDTGIHSREWITHTATGIWTANKI-----VSDYGKDRVLTDLNAMDIFIELVT 238  
QY 137 NGSYEGGAILPGFFMMVHSLFRGTAKLPLVEVKPADFVVGKDTENIRLGV-VNGS 191  
Db 239 N-----PDGFAFTHSMNRLWRKNK--SIRPGIFCIGVDLNRNWKSGFGGNGS 283

RESULT 6  
US-09-710-099-10  
; Sequence 10, Application US/09710099  
; Patent No. 6441154  
; GENERAL INFORMATION:  
; APPLICANT: Donoho, Gregory  
; APPLICANT: Hilbun, Erin  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Nehls, Michael  
; APPLICANT: Friedrich, Glenn









COUNTRY: U.S.A.  
ZIP: 07110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/459,065  
FILING DATE: 02-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/832,117  
FILING DATE: 06-FEB-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Roseman, Catherine R  
REGISTRATION NUMBER: 34,240  
REFERENCE/DOCKET NUMBER: 8589  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (201) 235-6208  
TELEFAX: (201) 235-3500  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3165 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Endothia parasitica (Cryphonectria  
ORGANISM: parasitica)  
STRAIN: EP713  
US-08-459-065-3

Query Match 6.4%; Score 81.5; DB 2; Length 3165;  
Best Local Similarity 20.3%; Pred. No. 20;  
Matches 45; Conservative 26; Mismatches 68; Indels 83; Gaps 7;  
QY 11 HSVFSITEDGKTFRRWRLSTGVFQTEDELSHLHPLLGDAMREIKGI-----GVAS 61  
Db 751 HDFRLIEITKTCRR-----NPPE-----NLQAKLEDTARKVCSVWQYNIMIASVAF 799  
QY 62 VVPTQNTVIERFSQKYFHISP-----IWKAKNGCVKWNKNPSEVGADRVANVVA 112  
Db 800 LVPLYFTLYVPYLQFYLVHVDPGDYILLPPVLWLVWTLNLCYGY-----ACDAWCRLFF 851  
QY 113 FVKEYGKNGII-----IDMGT 128  
b 852 FVEAGKKELVHSSEESSDPSSTLLIPTMGTRGDHVPFRFFANMAVLGVKTHLLKLOT 911  
QY 129 ATTVDLVVNGSYEGGAILPGFFMMVHSLFRG--TAKLPLVEV 168  
Db 912 ATYGLENLKKGLGSLLPGYLQNHYSVLRGYKAAFTPHVEL 953

Search completed: June 24, 2003, 21:49:22  
Job time : 8.18642 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:08:46 ; Search time 11.2158 Seconds  
(without alignments)  
2373.336 Million cell updates/sec

Title: US-09-813-453A-9  
Perfect score: 1281  
Sequence: 1 MYLLVDVGNTHSVFSITEDG.....HEIFDEDLTIKGVYHFCFGD 246

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues  
Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA.\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1281	100.0	246	9	US-09-813-453A-9
2	405.5	31.7	258	9	US-09-813-453A-2
3	399.5	31.2	262	9	US-09-813-453A-45
4	389.5	30.4	254	9	US-09-813-453A-47
5	375.5	29.3	258	9	US-09-813-453A-49
6	369.5	28.8	233	9	US-09-813-453A-17
7	361.5	28.2	255	9	US-09-813-453A-7
8	359.5	28.1	260	9	US-09-813-453A-51
9	342.5	26.7	265	9	US-09-813-453A-4
10	333.5	26.0	250	9	US-09-813-453A-3
11	321	25.1	256	9	US-09-813-453A-55
12	295.5	23.1	258	9	US-09-813-453A-6
13	286	22.3	272	9	US-09-813-453A-5
14	285	22.2	273	9	US-09-813-453A-10
15	282	22.0	272	9	US-09-712-363-276
16	279.5	21.8	219	9	US-09-813-453A-57
17	260	20.3	262	9	US-09-813-453A-8
18	229.5	17.9	241	9	US-09-813-453A-63
19	228.5	17.8	257	9	US-09-813-453A-53

20	221	17.3	212	9	US-09-813-453A-59	Sequence 59, Appl
21	203	15.8	262	9	US-09-813-453A-11	Sequence 11, Appl
22	193.5	15.1	229	9	US-09-813-453A-12	Sequence 12, Appl
23	165	12.9	223	9	US-09-895-913A-74	Sequence 14, Appl
24	165	12.9	223	9	US-09-813-453A-14	Sequence 14, Appl
25	165	12.9	223	9	US-09-813-453A-67	Sequence 67, Appl
26	162	12.6	244	9	US-09-813-453A-41	Sequence 41, Appl
27	152.5	11.9	460	9	US-09-813-453A-39	Sequence 39, Appl
28	152	11.9	267	9	US-09-813-453A-15	Sequence 15, Appl
29	150.5	11.7	592	9	US-09-813-453A-22	Sequence 22, Appl
30	146	11.4	592	9	US-09-813-453A-43	Sequence 43, Appl
31	145.5	11.4	249	9	US-09-813-453A-70	Sequence 70, Appl
32	145	11.3	248	9	US-09-813-453A-20	Sequence 20, Appl
33	139.5	10.9	209	9	US-09-813-453A-21	Sequence 21, Appl
34	138	10.8	257	9	US-09-813-453A-13	Sequence 13, Appl
35	117.5	9.2	249	9	US-09-813-453A-61	Sequence 61, Appl
36	103	8.0	242	9	US-09-813-453A-65	Sequence 65, Appl
37	97	7.6	317	12	US-10-043-238-1	Sequence 1, Appl
38	97	7.6	317	12	US-10-043-238-3	Sequence 3, Appl
39	90.5	7.1	351	9	US-10-200-910-2	Sequence 2, Appl
40	90.5	7.1	351	9	US-10-200-910-10	Sequence 10, Appl
41	90.5	7.1	436	9	US-10-076-535-2	Sequence 2, Appl
42	90.5	7.1	436	9	US-10-200-910-6	Sequence 6, Appl
43	87.5	6.8	541	10	US-09-873-409-7	Sequence 7, Appl
44	87.5	6.8	1058	10	US-09-873-409-4	Sequence 4, Appl
45	87.5	6.8	1222	10	US-09-873-409-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1

US-09-813-453A-9  
; Sequence 9, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 246  
; TYPE: PRT  
; ORGANISM: Thermotoga maritima  
US-09-813-453A-9

Query Match	100.0%	Score 1281;	DB 9;	Length 246;
Best Local Similarity	100.0%	Pred. No. 1.1e-118;		
Matches	246;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
QY	1	MYLLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDELSHLHPLLGDMREIKGIGVA	60	
Db	1	MYLLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDELSHLHPLLGDMREIKGIGVA	60	
QY	61	SVVPTQNTVIERFSQYFHSPIWVKAKNGCVKWNKPNPSEVGADRVANVAVFVKEYGKN	120	
Db	61	SVVPTQNTVIERFSQYFHSPIWVKAKNGCVKWNKPNPSEVGADRVANVAVFVKEYGKN	120	
QY	121	GIIDMGTTATVDLVVNGSYEGGAILPGFFMMVHSLFRGTAKLPLVEVKPADVVGKDTTE	180	
Db	121	GIIDMGTTATVDLVVNGSYEGGAILPGFFMMVHSLFRGTAKLPLVEVKPADVVGKDTTE	180	
QY	181	ENIRLGVVNGSVYALEGIIGRIKEYVDLPVVLTTGGQSKIVKDMIKHEIFDEDLTIKGVY	240	
Db	181	ENIRLGVVNGSVYALEGIIGRIKEYVDLPVVLTTGGQSKIVKDMIKHEIFDEDLTIKGVY	240	













GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:36:04 ; Search time 8.97264 Seconds  
(without alignments)  
2635.685 Million cell updates/sec

Title: US-09-813-453A-9  
Perfect score: 1281  
Sequence: 1 MYLLVDVGNTHSVFSITEDG.....HEIFDEDLTIKGVYHFCFGD 246

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1281	100.0	246	2 D72320	conserved hypothet
2	389.5	30.4	254	2 F83660	hypothetical prote
3	378.5	29.5	259	2 AF1102	conserved hypothet
4	377.5	29.5	259	2 AF1464	conserved hypothet
5	369.5	28.8	233	2 S66100	conserved hypothet
6	360.5	28.1	273	2 E97293	probable transcrip
7	359.5	28.1	261	2 B87489	transcription acti
8	342.5	26.7	265	2 T36391	hypothetical prote
9	289	22.6	274	2 H86937	conserved hypothet
10	286	22.3	272	2 A70955	hypothetical prote
11	285	22.2	273	2 D71326	conserved hypothet
12	260	20.3	262	2 E75516	conserved hypothet
13	204.5	16.0	224	2 A99571	conserved hypothet
14	203	15.8	262	2 F70165	conserved hypothet
15	193.5	15.1	229	2 E70465	hypothetical prote
16	179	14.0	223	2 G71887	hypothetical prote
17	165	12.9	223	2 F64627	hypothetical prote
18	152	11.9	267	2 I40327	baf protein - Bord
19	150.5	11.7	592	2 H82031	probable biotin-[a
20	146	11.4	592	2 B81009	BirA protein/Bvg a
21	145	11.3	248	2 H83111	hypothetical prote
22	139.5	10.9	209	2 H81382	hypothetical prote
23	138	10.8	257	2 S75559	hypothetical prote
24	124.5	9.7	276	2 AI2292	hypothetical prote
25	103	8.0	242	2 A82637	conserved hypothet
26	100.5	7.8	313	2 E97221	transcription regu
27	98	7.7	315	2 F70203	xylose operon regu
28	95	7.4	780	2 B84004	ATP-dependent DNA
29	94.5	7.4	287	2 C82158	ROK family protein

30	94	7.3	397	2 T11786	aspartate transami
31	94	7.3	1350	2 AF2005	RNA polymerase bet
32	93.5	7.3	717	2 B32838	DNA-directed RNA p
33	92.5	7.2	320	2 D95077	glucokinase [impor
34	92	7.2	330	1 H64404	hypothetical prote
35	91.5	7.1	436	2 F64362	3-phosphoshikimate
36	91.5	7.1	516	2 A96753	probable threonine
37	90.5	7.1	824	2 D70337	poly A polymerase
38	90	7.0	1470	2 S45323	genome polyprotein
39	89.5	7.0	325	2 H97944	glucokinase (EC 2.
40	89.5	7.0	702	2 A96959	oxygen-sensitive r
41	89	6.9	322	2 AC1242	glucose kinase hom
42	89	6.9	409	2 AI1857	tryptophan synthas
43	89	6.9	675	2 S53832	NADH2 dehydrogenas
44	89	6.9	2475	2 T00047	gellan lyase (EC 4
45	88.5	6.9	275	2 A69413	conserved hypothet

ALIGNMENTS

RESULT 1

D72320  
Conserved hypothetical protein - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: D72320  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; H Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genom  
A:Reference number: A72200; MUID:99287316; PMID:10360571  
A:Accession: D72320  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-246 <ARN>  
A:Cross-references: GB:AE001754; GB:AE000512; NID:g4981417; PIDN:AAD35964.1; PID:g4  
A:Experimental source: strain MSB8  
C:Genetics:  
A:Gene: TM0883  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match	100.0%	Score 1281;	DB 2;	Length 246;
Best Local Similarity	100.0%;	Pred. No. 4.3e-99;		
Matches 246;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MYLLVDVGNTHSVFSITEDGKTRRRLSTGVFQTEDELFSLHPLLLGDAMREIKGIVA	60	
Db	1	MYLLVDVGNTHSVFSITEDGKTRRRLSTGVFQTEDELFSLHPLLLGDAMREIKGIVA	60	
QY	61	SVVPTQNTVIERFSOKYFHISPIWVKAKNGCVKWNKPNSEVGADRVANVAVFVKEYGKN	120	
Db	61	SVVPTQNTVIERFSOKYFHISPIWVKAKNGCVKWNKPNSEVGADRVANVAVFVKEYGKN	120	
QY	121	GIIDMGTTATTVDLVNGSYEGGAILPGFFMMVHSLFRGTAKLPLVEVKPADFVVGKDT	180	
Db	121	GIIDMGTTATTVDLVNGSYEGGAILPGFFMMVHSLFRGTAKLPLVEVKPADFVVGKDT	180	
QY	181	ENIRLGVVNGSVVALEGIIGRIKEVYGDLPVVLTTGGQSKIVKDMIKHEIFDEDLTIKGVY	240	
Db	181	ENIRLGVVNGSVVALEGIIGRIKEVYGDLPVVLTTGGQSKIVKDMIKHEIFDEDLTIKGVY	240	
QY	241	HFCFGD 246		
Db	241	HFCFGD 246		

RESULT 2

F83660  
hypothetical protein BH0086 [imported] - Bacillus halodurans (strain C-125)  
C:Species: Bacillus halodurans  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001







RESULT 8  
T36391  
hypothetical protein SCE94.31c - Streptomyces coelicolor  
C;Species: Streptomyces coelicolor  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jun-2000  
C;Accession: T36391  
R;Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, April 1999  
A;Reference number: 221573  
A;Accession: T36391  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-265 <OLI>  
A;Cross-references: EMBL:AL049628; PIDN:CAB40880.1; GSPDB:GN00070; SCOEDB:SCE94.31c  
A;Experimental source: strain A3(2)  
C;Genetics:  
A;Gene: SCOEDB:SCE94.31c  
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 26.7%; Score 342.5; DB 2; Length 265;  
Best Local Similarity 35.7%; Pred. No. 4.9e-21;  
Matches 92; Conservative 45; Mismatches 98; Indels 23; Gaps 9;  
QY 1 MYLLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDEL-----FSLHPLLLGDAMRE- 53  
| | :|||:| : : ||:| :| ||| :||| :| :  
Db 1 MLTTIDVGNTHTVLGLFGEDIVEHWRISTDSRRRTADELAVLQGLMGHPLGLGDELGDG 60  
| | :|||:| : : ||:| :| ||| :||| :| :  
QY 54 IKGIGVASVVPQTNTVIERFSQKYFHISPIWVKAKGCVKWNV---KNPSEVGADRVAN 109  
| | :|||:| : : ||:| :| ||| :||| :| :  
Db 61 IDGIAICATVPSVLHELREVTTRYGDVPA-VLVEPG-VKTGVPILTDPKKEVGADRIIN 118  
| | :|||:| : : ||:| :| ||| :||| :| :  
QY 110 VVAFVKEYGKNGIIDMGTTATVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKPLVEV 168  
| | :|||:| : : ||:| :| ||| :||| :| :  
Db 119 AVAAVELYGGPAIVVDFGTATTFDAVSARGEYIGGVITAPGIEISVEALGVKGAQLRKIEV 178  
| | :|||:| : : ||:| :| ||| :||| :| :  
QY 169 KPADFVVGKDTENIRLGVVNGSVYALEGIIIGRIKEYGDLP----VVLTTGGQSKIIVKDM 224  
| | :|||:| : : ||:| :| ||| :||| :| :  
Db 179 ARPRSVIGKNTVEAMQSGIYVGFAGQVDGVVNRMARELADDDVTVIATGGLAPMV--L 236  
| | :|||:| : : ||:| :| ||| :||| :| :  
QY 225 IKHEIFDE--LTIKGV 239  
| | :|||:| : : ||:| :| ||| :||| :| :  
Db 237 GESSVIDEHEPWLTLMGL 254  
| | :|||:| : : ||:| :| ||| :||| :| :

RESULT 9  
H86937  
conserved hypothetical protein ML0232 [imported] - Mycobacterium leprae  
Species: Mycobacterium leprae  
C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
C;Accession: H86937  
R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Holroyd, R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd, M.A.; Rutherford, K.M.  
Nature 409, 1007-1011, 2001  
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Squares, R.; Title: Massive gene decay in the leprosy bacillus.  
A;Reference number: A86909; MUID:21128732; PMID:11234002  
A;Accession: H86937  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-274 <STO>  
A;Cross-references: GB:AL450380; NID:g13092576; PIDN:CAC29740.1; GSPDB:GN00147  
C;Genetics:  
A;Gene: ML0232  
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 22.6%; Score 289; DB 2; Length 274;  
Best Local Similarity 32.3%; Pred. No. 1.4e-16;  
Matches 83; Conservative 47; Mismatches 107; Indels 20; Gaps 8;  
QY 1 MYLLVDVGNTHSVFSI----TEDGKTFRRWRLSTGVFQTEDELFSHLHPLLLGDAMREIKG 56

Db 1 MLLAIDVRNTHTVVGLSGSKEHAKVVQWRIrTSEVTADELALIIDGLIGDDSERLAG 60  
QY 57 IGVASVVPQTNTVIERFSQKYFHISP---IWKAKNGCVKWNKPNPSEVGADRVANVAF 113  
| | :|||:| : : ||:| :| ||| :||| :| :  
Db 61 AAALSTVPVSLHEVRIMLDQYWPSVPHVLIEPGVRTG-IPLLVDNPKKEVGADRVNCLAA 119  
| | :|||:| : : ||:| :| ||| :||| :| :  
QY 114 VKEYGKNGIIDMGTTATVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKPLVEVKPAD 172  
| | :|||:| : : ||:| :| ||| :||| :| :  
Db 120 FHKFGQAAIIVDFGSSICVDVVSAGGEFLGGAIAIPGVQVSSDAAAARSALRRVELARPR 179  
| | :|||:| : : ||:| :| ||| :||| :| :  
QY 173 FVVGKDTENIRLGVVNGSVYALEGIIGR----IKEYVGDL----PVVLTGGQSKIIVKDM 224  
| | :|||:| : : ||:| :| ||| :||| :| :  
Db 180 SVVGKNTVECMQAGVVFAGLVLDGLVGRMQDVVEFSGDLGNRVAVVATGHTAPLLLPE 239  
| | :|||:| : : ||:| :| ||| :||| :| :  
QY 225 IKHEI--FDEDLTIKGV 239  
| | :|||:| : : ||:| :| ||| :||| :| :  
Db 240 L-HTVDHYDRHLTLHGL 255  
| | :|||:| : : ||:| :| ||| :||| :| :

RESULT 10  
A70955  
hypothetical protein Rv3600c - Mycobacterium tuberculosis (strain H37RV)  
C;Species: Mycobacterium tuberculosis  
C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C;Accession: A70955  
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordo  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete gen  
A;Reference number: A70500; MUID:98295987; PMID:9634230  
A;Accession: A70955  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-272 <COL>  
A;Cross-references: GB:Z95557; GB:AL123456; NID:g3242276; PIDN:CAB08944.1; PID:g2113  
A;Experimental source: strain H37RV  
C;Genetics:  
A;Gene: Rv3600c  
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 22.3%; Score 286; DB 2; Length 272;  
Best Local Similarity 31.4%; Pred. No. 2.5e-16;  
Matches 80; Conservative 49; Mismatches 108; Indels 18; Gaps 7;  
QY 1 MYLLVDVGNTHSVFSI----TEDGKTFRRWRLSTGVFQTEDELFSHLHPLLLGDAMREIKG 56  
| | :|||:| : : ||:| :| ||| :||| :| :  
Db 1 MLLAIDVRNTHTVVGLSGMKEHAKVVQWRIrTSEVTADELALTIDGLIGEDSERLTG 60  
| | :|||:| : : ||:| :| ||| :||| :| :  
QY 57 IGVASVVPQTNTVIERFSQKYFHISP---IWKAKNGCVKWNKPNPSEVGADRVANVAF 113  
| | :|||:| : : ||:| :| ||| :||| :| :  
Db 61 TAALSTVPVSLHEVRIMLDQYWPSVPHVLIEPGVRTG-IPLLVDNPKKEVGADRVNCLAA 119  
| | :|||:| : : ||:| :| ||| :||| :| :  
QY 114 VKEYGKNGIIDMGTTATVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKPLVEVKPAD 172  
| | :|||:| : : ||:| :| ||| :||| :| :  
Db 120 YDRFRKAAIIVDFGSSICVDVVSAGGEFLGGAIAIPGVQVSSDAAAARSALRRVELARPR 179  
| | :|||:| : : ||:| :| ||| :||| :| :  
QY 173 FVVGKDTENIRLGVVNGSVYALEGIIGRIKE-----VYGDLPVVLTTGGQSKIIVKDMIK 226  
| | :|||:| : : ||:| :| ||| :||| :| :  
Db 180 SVVGKNTVECMQAGAVVFAGLVLDGLVGRIREDDVSGFSDHDVAIVATGHTAPLLLPEL- 238  
| | :|||:| : : ||:| :| ||| :||| :| :  
QY 227 HEI--FDEDLTIKGV 239  
| | :|||:| : : ||:| :| ||| :||| :| :  
Db 239 HTVDHYDQHLLTLQGL 253  
| | :|||:| : : ||:| :| ||| :||| :| :

RESULT 11  
D71326  
conserved hypothetical protein TP0431 - syphilis spirochete  
C;Species: Treponema pallidum subsp. pallidum (syphilis spirochete)  
C;Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 20-Jun-2000







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:26 ; Search time 25.2563 Seconds  
(without alignments)  
1297.879 Million cell updates/sec

Title: US-09-813-453A-9  
Perfect score: 1281  
Sequence: 1 MYLLVDVGNTHSVFSITEDG.....HEIFDEDLTIKGVHFCFGD 246

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

arched: 908470 seqs, 133250620 residues  
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1281	100.0	246	23	AAU91156	Thermotoga maritima
2	405.5	31.7	258	22	AAU01243	B. subtilis novel
3	405.5	31.7	258	23	AAU91149	Bacillus subtilis
4	399.5	31.2	262	23	AAU91170	Pantothenate kinas
5	389.5	30.4	254	23	AAU91171	Pantothenate kinas
6	378.5	29.5	259	23	ABB47661	Listeria monocytog
7	375.5	29.3	258	23	AAU91172	Pantothenate kinas
8	369.5	28.8	233	23	AAU91163	Pantothenate kinas
9	361.5	28.2	255	23	AAU91154	Geobacter sulfurre
10	359.5	28.1	260	23	AAU91173	Pantothenate kinas

11	342.5	26.7	265	23	AAU91151	Streptomyces coeli
12	333.5	26.0	250	23	AAU91150	Clostridium acetob
13	321	25.1	256	23	AAU91175	Pantothenate kinas
14	295.5	23.1	258	23	AAU91153	Rhodobacter capsul
15	286	22.3	272	23	AAU91152	Mycobacterium tube
16	285	22.2	273	23	AAU91157	Treponema pallidum
17	282	22.0	272	22	AAG81225	Mycobacterium tube
18	279.5	21.8	219	23	AAU91176	Pantothenate kinas
19	260	20.3	262	23	AAU91155	Deinococcus radiop
20	229.5	17.9	241	23	AAU91179	Pantothenate kinas
21	228.5	17.8	257	23	AAU91174	Pantothenate kinas
22	221	17.3	212	23	AAU91177	Pantothenate kinas
23	203	15.8	262	23	AAU91158	Borrelia burgdorfe
24	193.5	15.1	229	23	AAU91159	Aquifex aeolicus p
25	165	12.9	223	19	AAW98422	H. pylori GHPO 344
26	165	12.9	223	23	AAU91161	Helicobacter pylor
27	165	12.9	223	23	AAU91181	Pantothenate kinas
28	162	12.6	244	23	AAU91168	Pantothenate kinas
29	152.5	11.9	455	20	AAV38617	Neisseria gonorrh
30	152.5	11.9	455	21	AAV74908	Neisseria gonorrh
31	152.5	11.9	460	23	AAU91167	Pantothenate kinas
32	152.5	11.9	592	20	AAV38618	Neisseria gonorrh
33	152.5	11.9	592	21	AAV74911	Neisseria gonorrh
34	152	11.9	267	23	AAU91162	Bordella pertussis
35	150.5	11.7	455	21	AAV74910	Neisseria meningit
36	150.5	11.7	592	20	AAV38616	Neisseria meningit
37	150.5	11.7	592	21	AAV74913	Neisseria meningit
38	150.5	11.7	592	23	AAU91166	Pantothenate kinas
39	149.5	11.7	389	21	AAV74909	Neisseria meningit
40	146	11.4	592	20	AAV38615	Neisseria meningit
41	146	11.4	592	21	AAV74912	Neisseria meningit
42	146	11.4	592	23	AAU91169	Pantothenate kinas
43	145.5	11.4	249	23	AAU91182	Pantothenate kinas
44	145	11.3	248	23	AAU91164	Pantothenate kinas
45	139.5	10.9	209	23	AAU91165	Pantothenate kinas

ALIGNMENTS

RESULT 1  
AAU91156  
ID AAU91156 standard; Protein; 246 AA.  
XX  
AC AAU91156;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Thermotoga maritima pantothenate kinase Coax.  
XX  
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX  
OS Thermotoga maritima.  
XX  
PN WO200216601-A2.  
XX  
PD 28-FEB-2002.  
XX  
PF 24-AUG-2001; 2001WO-US26531.  
XX  
PR 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX  
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX  
PI Yocum RR, Patterson TA;  
XX  
DR WPI; 2002-269358/31.  
XX  
PT Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein

XX PS Claim 10; Page 74-75; 128pp; English.

XX CC The invention describes assays for identifying a (potential) antibiotic

CC comprising contacting an assay composition comprising a pantothenate

CC kinase (Coax) protein with a test compound, and determining the ability

CC of the test compound to inhibit the activity of the Coax protein, an

CC essential enzyme for the production of coenzyme A. Coax protein is a

CC valuable target for identifying bactericidal compounds. Coax modulating

CC agents can be used in an infectious animal model to determine the

CC efficacy, toxicity, or side effects of treatment with such an agent. This

CC is the amino acid sequence of a pantothenate kinase (Coax) protein

CC described in the invention.

XX SQ Sequence 246 AA;

Query Match 100.0%; Score 1281; DB 23; Length 246;

Best Local Similarity 100.0%; Pred. No. 7.9e-133;

Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MYLLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDELFSLHPLLLGDAMREIKGIGVA 60

1 MYLLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDELFSLHPLLLGDAMREIKGIGVA 60

61 SVVPTQNTVIERFSQKYFHISPIWVKAKNGCVKWNKPNPSEVGADRVANVYAFVKEYGKN 120

61 SVVPTQNTVIERFSQKYFHISPIWVKAKNGCVKWNKPNPSEVGADRVANVYAFVKEYGKN 120

121 GIIDMGATTVDLVVNGSYEGGAILPGFFMMVHSLFRGTAKPLPVEVKPADPVGKDT 180

121 GIIDMGATTVDLVVNGSYEGGAILPGFFMMVHSLFRGTAKPLPVEVKPADPVGKDT 180

181 ENIRLGVVNGSVVALEGIIGRIKEVYGDLPVLTGGQSKIVKDMIKHEIFDEDLTIKGVY 240

181 ENIRLGVVNGSVVALEGIIGRIKEVYGDLPVLTGGQSKIVKDMIKHEIFDEDLTIKGVY 240

241 HFCFGD 246

241 HFCFGD 246

RESULT 2

AAU01243

ID AAU01243 standard; Protein; 258 AA.

XX AC AAU01243;

XX DT 18-JUL-2001 (first entry)

XX B. subtilis novel pantothenate kinase encoded by the gene coax.

KW Pantothenate kinase; coax; pantothenate biosynthesis; vitamin B5;

KW nutritional supplement; panto-compound; pantoate.

XX OS Bacillus subtilis.

XX PN WO200121772-A2.

XX 29-MAR-2001.

XX 21-SEP-2000; 2000WO-US25993.

XX 21-SEP-1999; 99US-0400494.

PR 07-JUN-2000; 2000US-0210072.

PR 28-JUL-2000; 2000US-0221836.

PR 24-AUG-2000; 2000US-0227860.

XX (OMNI-) OMNIGENE BIOPRODUCTS.

PA Yocum RR, Patterson TA, Hermann T, Pero JG;

XX WPI; 2001-218644/22.

DR N-PSDB; AAS00984.

XX PT New recombinant microorganism which overexpress a Bacillus subtilis

PT pantothenate biosynthetic enzyme, useful for the high yield production

PT of panto-compounds such as pantothenate and pantoate -

XX Example 14; Fig 23; 292pp; English.

XX The sequence represents a novel B. subtilis pantothenate kinase (encoded

CC by gene coax), an enzyme of the pantothenate biosynthetic pathway.

CC Pantothenate, also known as vitamin B5, is used as a nutritional

CC supplement in mammals and humans. The invention concerns methods of

CC producing recombinant microorganisms overexpressing at least one Bacillus

CC subtilis pantothenate biosynthetic enzyme. The microorganisms and methods

CC of producing them are useful for producing a panto-compound such as

CC pantothenate or pantoate, which is a nutritional requirement for

CC livestock and humans. The methods are also useful for the identification

CC of pantothenate kinase modulators. Panto-compounds are produced at a

CC significantly higher yield than prior art methods and can be produced

CC independent of the need to feed precursors which decreases expense.

XX SQ Sequence 258 AA;

Query Match 31.7%; Score 405.5; DB 22; Length 258;

Best Local Similarity 39.2%; Pred. No. 4e-36;

Matches 100; Conservative 42; Mismatches 88; Indels 25; Gaps 9;

1 MYLLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDE-----LFSLHPLLLGDAMRE 53

1 LLLVIDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDEFGMLRLSLFDH---SGLMFEQ 56

54 IKGIGVASVVPTQNTVIERFSQKYFHISPIWVKAKNGCVKWNK--NPSEVGADRVAN 109

57 IDGIISSVVPPIFMALERMCTKYFHIEPQIVGPGMKTGL---NIKYDNPKEVGADRVN 113

110 VVAFVKEYGKNGIIDMGATTVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKPLPVEV 168

114 AVAAIHLYGNPLIVVDFGATTTCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIE 173

169 KPADFVVGKDTENIRLGVVNGSVVALEGIIGRIK-EVYGDLPVLTGGQSKIV---KDM 224

174 TRPDNIIGKNTVSAMQSGILFGYGVQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDC 233

225 IKHEIFDEDLTIKGV 239

234 I--DIVDPFLTLKGL 246

RESULT 3

AAU91149

ID AAU91149 standard; Protein; 258 AA.

XX AC AAU91149;

XX 05-JUN-2002 (first entry)

XX Bacillus subtilis pantothenate kinase Coax.

XX Pantothenate kinase; Coax; antibiotic; antimicrobial;

KW pantothenate kinase modulator; coenzyme A; bactericidal compound.

XX OS Bacillus subtilis.

XX WO200216601-A2.

XX 28-FEB-2002.

XX 24-AUG-2001; 2001WO-US26531.

XX 24-AUG-2000; 2000US-227860P.

PR 20-MAR-2001; 2001US-0813453.

XX (OMNI-) OMNIGENE BIOPRODUCTS INC.



```
PR 20-MAR-2001; 2001US-0813453.
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
PA Yocum RR, Patterson TA;
XX WPI; 2002-269358/31.
DR N-PSDB; ABK54192.
XX Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein -
XX Claim 10; Page 100; 128pp; English.
XX The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX Sequence 254 AA;
SQ
Query Match 30.4%; Score 389.5; DB 23; Length 254;
Best Local Similarity 35.6%; Pred. NO. 2.3e-34;
Matches 90; Conservative 54; Mismatches 88; Indels 21; Gaps 8;
QY 1 MYLLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDE-----LFSHLHPLLGDMRE 53
Db 1 MILVIDVGNNTNVLGVYQDETLLVHHWRLATSRQKTEDEYAMTVRSLEFDH----AGLQFQD 56
QY 54 IKGIGVASVVPQTNTVIERFSQKYFHISPIWV--KAKNGCVKWNVK--NPSEVGADRVAN 109
Db 57 IDGIVISSVVPMPMFSLQMKCKKYFHTVPMIIGPIKGTGL---NIKYDNPKEVGADRVIN 113
QY 110 VVAFVKEYGKNGIIDMGTTATVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKLPLVEV 168
Db 114 AVAAIELYGYPAIVDFGTATTYCLINEKKOYAGGVIAPIGIMISTEALYHRASKLPRIE 173
QY 169 KPADFVVGKDTTEENIRLGVVNGSVYALEGIIIGRIK-EVYGDLPVVLTTGGOSKIV-KDMIK 226
Db 174 AKPKQVVGNTIDSMQSGIFYGYVSQDGVVVKRMKAQAESEPKVIATGGLAKLIGTSET 233
QY 227 HEIFDEDLTIKGV 239
Db 234 IDVIDSFLTLKGL 246
RESULT 6
ID ABB47661 standard; Protein; 259 AA.
XX ABB47661
AC ABB47661;
XX 05-FEB-2002 (first entry)
XX Listeria monocytogenes protein #365.
DE Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW vitamin B12; bacterial infection; disease.
XX Listeria monocytogenes.
OS
XX WO200177335-A2.
PN 18-OCT-2001.
PD 11-APR-2001; 2001WO-FR01118.
PF
```

```
XX 11-APR-2000; 2000FR-0004629.
XX (INSP ) INST PASTEUR.
PA Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
XX Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
PI Rose M, Voss H;
XX WPI; 2002-010914/01.
DR Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
XX and prevention of Listeria and related bacterial infections, and
XX related polypeptides -
XX Claim 6; SEQ ID No 366; 192pp; French.
XX The present invention relates to the genome sequence of Listeria
CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
CC it are useful for selecting probes and primers for detecting genes in L.
CC monocytogenes and related organisms, and for studying genetic
CC polymorphisms and other genomes. The present sequence is a protein
CC encoded by the genome sequence of the present invention. Proteins
CC expressed from the genome sequence are useful for raising specific
CC antibodies, identification of L. monocytogenes and related organisms, and
CC for biosynthesis and biodegradation, especially biosynthesis of vitamin
CC B12. The genome sequence and proteins encoded by it are also useful for
CC selecting compounds that regulate gene expression and cell replication
CC and modulate L. monocytogenes-related diseases. In addition, the genome
CC sequence and proteins encoded by it are useful in pharmaceutical and
CC vaccine compositions for the treatment or prevention of infections by L.
CC monocytogenes and related organisms.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 259 AA;
SQ
Query Match 29.5%; Score 378.5; DB 23; Length 259;
Best Local Similarity 35.6%; Pred. No. 3.8e-33;
Matches 90; Conservative 52; Mismatches 90; Indels 21; Gaps 7;
QY 1 MYLLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDEL-----FSLHPLLGDM 51
Db 1 MILVIDVGNNTNCTGVYEKQKLLKHWMTTDRHRTSDELGMTVLNFFSYANLTP----- 54
QY 52 REIKGIGVASVVPQTNTVIERFSQKYFHISPIWV--KAKNGCVKWNKNPSEVGADRVAN 109
Db 55 SDIQGIISVVPPIHMETMVCVRYFNIRPLIVGPGIKTG-LNLKVDNPREIGSDRVN 113
QY 110 VVAFVKEYGKNGIIDMGTTATVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKLPLVEV 168
Db 114 AVAASEEYGTPIVVDFGTATTCYIDESGVYQGGAIAPGIMISTEALYNRAAKLPRVDI 173
QY 169 KPADFVVGKDTTEENIRLGVVNGSVYALEGIIIGRIKEVYGDLPVVL-TGGOSKIVKDMIKH 227
Db 174 AESSQIIGKSTVSSMQAGIFYGVFGVQCCEGIIAEMKKQSNASPVVVATGGLARMIKSSA 233
QY 228 -EIFDEDLTIKGV 239
Db 234 VDILDPLTLKGL 246
RESULT 7
ID AAU91172 standard; Protein; 258 AA.
XX AAU91172;
AC
```



```
XX DT 05-JUN-2002 (first entry)
XX DE Pantothenate kinase (Coax) #10.
XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX OS Bacillus stearothermophilus.
XX PN WO200216601-A2.
XX PD 28-FEB-2002.
XX PF 24-AUG-2001; 2001WO-US26531.
XX PR 24-AUG-2000; 2000US-227860P.
XX PR 20-MAR-2001; 2001US-0813453.
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
PI Yocum RR, Patterson TA;
XX WPI; 2002-269358/31.
DR N-PSDB; ABK54193.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein -
XX Claim 10; Page 101-102; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX
SQ Sequence 258 AA;
Query Match 29.3%; Score 375.5; DB 23; Length 258;
Best Local Similarity 37.5%; Pred. No. 8.2e-33;
Matches 94; Conservative 48; Mismatches 92; Indels 17; Gaps 9;
QY 1 MYLLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDELFSLHPL--GDAMREIKGI 57
Db 1 MIFVLDVGNTNTVLGVYDGDGLKHHWRIETSRKTEDEYGMMLKALLNHVGLQFSDIRGI 60
QY 58 GVASVPTQNTVIERFSQKYFHISPIWV--KAKNGC-VKWNVKNPSEVGADRVANVAVFV 114
Db 61 IISVVPPIMFALERMCLKYFHKPLIVGPGIKTGLDIKYD--NPREVGADRVNAVAGI 118
QY 115 KEYCKNGIIDMGTTATVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKLPLVEVKPADF 173
Db 119 HLYGSPLIIVDFGTATTTCYINEHKQYMGGAIAPIGIMISTEALFARAAKLPRIEIRPDD 178
QY 174 VVGKDTENIRLGVVNGSVYALEGIGRIKEVYGDLP--VVLTGQSQKIV---KDMIKHE 228
Db 179 IIGKNTVSAMQAGILYGYVGQVEGIVSRMK-AKSKIPPKVIATGGLAPLIASESDI--D 235
QY 229 IFDEDLTIKGV 239
Db 236 VVDPFLTLTGL 246
```

RESULT 8  
AAU91163  
ID AAU91163 standard; Protein; 233 AA.

```
XX AC AAU91163;
XX DT 05-JUN-2002 (first entry)
XX DE Pantothenate kinase (Coax) #1.
XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX OS Bacillus subtilis.
XX PN WO200216601-A2.
XX PD 28-FEB-2002.
XX PF 24-AUG-2001; 2001WO-US26531.
XX PR 24-AUG-2000; 2000US-227860P.
XX PR 20-MAR-2001; 2001US-0813453.
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
PI Yocum FR, Patterson TA;
XX WPI; 2002-269358/31.
DR N-PSDB; ABK54169.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein -
XX Disclosure; Page 81-82; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX
SQ Sequence 233 AA;
Query Match 28.8%; Score 369.5; DB 23; Length 233;
Best Local Similarity 40.5%; Pred. No. 3.2e-32;
Matches 87; Conservative 33; Mismatches 76; Indels 19; Gaps 6;
QY 1 MYLLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDE-----LFSLHPLLGDAMRE 53
Db 1 MLLVLDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMLRLSLFDH----SGLMFEQ 56
QY 54 IKGIGVASVPTQNTVIERFSQKYFHISPP--IWKAKNGCVKWNVK--NPSEVGADRVAN 109
Db 57 IDGIISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGL---NIKYDNPKEVGADRVN 113
QY 110 VVAFVKEYGKNGIIDMGTTATVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKLPLVEV 168
Db 114 AVAAIHLVGNPLIVDFGTATTTCYIDENKQYMGGAIAPIGITSTEALYSRAAKLPRIE 173
QY 159 KPADFVVGKDTENIRLGVVNGSVYALEGIGRIK 203
Db 174 TRPDNIIGKNTVSAMQSGILFGYVGQVEGIVKRMK 208
```

RESULT 9  
AAU91154  
ID AAU91154 standard; Protein; 255 AA.  
XX  
AC AAU91154;

XX 05-JUN-2002 (first entry)  
DT Geobacter sulfurreducens pantothenate kinase Coax.  
XX  
DE Pantothenate kinase; Coax; antibiotic; antimicrobial;  
XX pantothenate kinase modulator; coenzyme A; bactericidal compound.  
KW Geobacter sulfurreducens.  
XX WO200216601-A2.  
XX 28-FEB-2002.  
XX 24-AUG-2001; 2001WO-US26531.  
XX 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
PA Yocum RR, Patterson TA;  
XX WPI; 2002-269358/31.  
DR Identifying potential antibiotic or antimicrobial agent, comprises  
XX contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -  
PT  
XX Claim 10; Page 72-73; 128pp; English.  
PS The invention describes assays for identifying a (potential) antibiotic  
XX comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.  
XX  
SQ Sequence 255 AA;  
Query Match 28.2%; Score 361.5; DB 23; Length 255;  
Best Local Similarity 36.8%; Pred. No. 2.8e-31;  
Matches 91; Conservative 49; Mismatches 98; Indels 9; Gaps 6;  
1 MYLLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDE--LFSHLHPLLGDMREIKGI 57  
1 MLLVIDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDE--LFSHLHPLLGDMREIKGI 57  
58 GVASVPTQNTVIERFSQKYPHISPIW--KAKNGCVKWNKPNSEVGADRVANVAVFK 115  
61 IISVVPPLTGVLERLSLGYFGMRPLVVGPGIKTG-MPIQYDNPREVGADRVNVAAGYE 119  
116 EYKNGIIDMGTAATVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKLPLVEVKPADFV 174  
120 KYRTSLIIVDFGTATFDYVNRKGEYCGGAIAPGLVTSTEALFQRASKLPRVDIIRPSAI 179  
175 VGKDTENIRLGVNVSVALEGIIGRIKEVYGDLP-VVLTGGQSKIVKDMIKH-EIFDE 232  
180 IARTVNSMQAGIYGVGLVDEIVTRMKAESKDAPRVATGGLASLIAPESKTEAVEE 239  
233 DLTIKGV 239  
240 YLTLEGL 246  
RESULT 10  
AAU91173  
ID AAU91173 standard; Protein; 260 AA.  
XX

AC AAU91173;  
XX 05-JUN-2002 (first entry)  
DT Pantothenate kinase (Coax) #11.  
XX Pantothenate kinase; Coax; antibiotic; antimicrobial;  
DE pantothenate kinase modulator; coenzyme A; bactericidal compound.  
KW Geobacter crescentus.  
XX WO200216601-A2.  
XX 28-FEB-2002.  
XX 24-AUG-2001; 2001WO-US26531.  
XX 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
PA Yocum RR, Patterson TA;  
XX WPI; 2002-269358/31.  
DR N-PSDB; ABK54194.  
XX Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -  
XX  
XX Claim 10; Page 102-103; 128pp; English.  
PS The invention describes assays for identifying a (potential) antibiotic  
XX comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.  
XX  
SQ Sequence 260 AA;  
Query Match 28.1%; Score 359.5; DB 23; Length 260;  
Best Local Similarity 35.5%; Pred. No. 4.8e-31;  
Matches 89; Conservative 53; Mismatches 94; Indels 15; Gaps 9;  
1 MYLLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDELFSLHPLL--GDAMREIKGI 57  
1 MLLAIEQGNTNTMFAIHDGASVVAQWRSATESRTRTADEYVVVWLSQLLSMQGLGFRAIDAV 60  
58 GVASVPTQNTVIERFSQKYPHISPIW--KAKNGCVKWNKPNSEVGADRVANVAVFK 115  
61 IISVVPQSIFNRLNLSRRYFNVEPLVIGENAKLG-IDVRIEKPSEAGADRLVNAIGAAM 119  
116 EYKNGIIDMGTAATVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKLPLVEV-KPA-D 172  
120 VYPGPLVVIDSGTATFDIIVADGAFEGGIIAPGINLSMQALHEAAKLPRIAIORPAGN 179  
173 FVVGKDTENIRLGVNVSVALEGIIGRIKEVYGD-LPVVLTGGQSKIVK--DMIKHE 228  
180 RIVGTDVTSAMQSGVFWGYISLIEGLVARIKAEERGEPTMTVIATGGVASLFEFEGATSDIH- 238  
229 IFDEDLTIKGV 239  
239 -FDSDLTIRGL 248  
RESULT 11  
AAU91151

ID AAU91151 standard; Protein; 265 AA.  
XX AC AAU91151;  
XX DT 05-JUN-2002 (first entry)  
XX DE Streptomyces coelicolor pantothenate kinase Coax.  
XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
XX KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX OS Streptomyces coelicolor.  
XX PN WO200216601-A2.  
XX PD 28-FEB-2002.  
XX PF 24-AUG-2001; 2001WO-US26531.  
XX PR 24-AUG-2000; 2000US-227860P.  
XX PR 20-MAR-2001; 2001US-0813453.  
XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX PI Yocum RR, Patterson TA;  
XX DR WPI; 2002-269358/31.  
XX

Identifying potential antibiotic or antimicrobial agent, comprises contacting composition comprising pantothenate kinase (Coax) protein with test compound and identifying inhibitor of the Coax protein.  
Claim 10; Page 69-70; 128pp; English.  
The invention describes assays for identifying a (potential) antibiotic comprising contacting an assay composition comprising a pantothenate kinase (Coax) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulating agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (Coax) protein described in the invention.

XX SQ Sequence 265 AA;

Query Match 26.7%; Score 342.5; DB 23; Length 265;  
Best Local Similarity 35.7%; Pred. NO. 3.7e-29;  
Matches 92; Conservative 45; Mismatches 98; Indels 23; Gaps 9;

QY 1 MYLLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDEL-----FSLHPLLLGDAMRE- 53  
Db 1 MLLTIDVGNTHTVLGLFDGEDIVEHWRISTDSRRTADELAVLLQGLMGHPLLLGDELGDG 60  
QY 54 IKGIGVASVVTQNTVIERFSQKYFHISPIWVKAKNGCVKWNV---KNPSEVGADRVAN 109  
Db 61 IDGIAICATVPSVLHELREVTTRYGVDVPA-VLVEPG-VKTGVPILTDHPKEVGADRIIN 118  
QY 110 VVAFVKEYKNGIIDIIMGTATTVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKLPLVEV 168  
Db 119 AVAAVELYGGPAIVVDFGTATTFDVAVSARGEYIGGVIAPIGELISVEALGVKGAQLRKIEV 178  
QY 169 KPADFVVGKDTENIRLVNGSVYALEGIIGRIKEVYGDLP----VLTGGQSKIVKDM 224  
Db 179 ARPRSVIGKNTVEAMQSGIVYGFAGQVDGVNRMARELADDDVTVIATGGLAPMV--L 236  
QY 225 IKHEIFDED---LTIKV 239  
Db 237 GESSVIDEHEPWLTLML 254

RESULT 12

AAU91150  
ID AAU91150 standard; Protein; 250 AA.  
XX AC AAU91150;  
XX DT 05-JUN-2002 (first entry)  
XX DE Clostridium acetobutylicum pantothenate kinase Coax.  
XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
XX KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX OS Clostridium acetobutylicum.  
XX PN WO200216601-A2.  
XX PD 28-FEB-2002.  
XX PF 24-AUG-2001; 2001WO-US26531.  
XX PR 24-AUG-2000; 2000US-227860P.  
XX PR 20-MAR-2001; 2001US-0813453.  
XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX PI Yocum RR, Patterson TA;  
XX DR WPI; 2002-269358/31.  
XX

Identifying potential antibiotic or antimicrobial agent, comprises contacting composition comprising pantothenate kinase (Coax) protein with test compound and identifying inhibitor of the Coax protein.  
Claim 10; Page 68-69; 128pp; English.

The invention describes assays for identifying a (potential) antibiotic comprising contacting an assay composition comprising a pantothenate kinase (Coax) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulating agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (Coax) protein described in the invention.

XX SQ Sequence 250 AA;

Query Match 26.0%; Score 333.5; DB 23; Length 250;  
Best Local Similarity 39.1%; Pred. NO. 3.4e-28;  
Matches 95; Conservative 39; Mismatches 72; Indels 37; Gaps 11;

QY 3 LLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDE-----LFSH--LHPLLGDAMRE 53  
Db 20 LVLDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDE-----LFSH--LHPLLGDAMRE 73  
QY 54 IKGIGVASVVTQNTVIERFSQKYFHISPIWV--KAKNGCVKWNV--NPSEVGADRVAN 109  
Db 74 VEGVISSVVPNIMYSLEHMIRKFKINPLVVGPIKTI---NIKYDNPKEVGADRVN 130  
QY 110 VVAFVKEYKNGIIDIIMGTATTVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKLPLVE- 167  
Db 131 AVAAHEIYKRSIIIDFGTATTFCVAVRENGDYLGAICPGIKVSSEALFEKAKLPRVEL 190  
QY 168 VKPADFVVGKDTENIRLVNGSVYALEGIIGRIKEVYGDLP-----VLTGGQSK 218  
Db 191 IKPA-YAICKNTISSIQSGIVRYLRQVKYLFKEKLE---NLPDGRRTRTSLVLTATGGLA 246  
QY 219 KIV 221  
Db 247 KLI 249

RESULT 13  
AAU91175  
ID AAU91175 standard; Protein; 256 AA.  
XX  
AC AAU91175;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Pantothenate kinase (Coax) #13.  
XX  
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX  
OS Clostridium difficile.  
XX  
PN WO200216601-A2.  
XX  
PD 28-FEB-2002.  
XX  
PF 24-AUG-2001; 2001WO-US26531.  
XX  
PA 24-AUG-2000; 2000US-227860P.  
XX 20-MAR-2001; 2001US-0813453.  
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
PI Yocum RR, Patterson TA;  
DR WPI; 2002-269358/31.  
DR N-PSDB; ABK54196.  
XX  
PT Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -  
XX  
PS Claim 6; Page 105; 128pp; English.  
XX  
CC The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.  
XX  
SQ Sequence 256 AA;  
Query Match 25.1%; Score 321; DB 23; Length 256;  
Best Local Similarity 34.4%; Pred. No. 8.4e-27;  
Matches 86; Conservative 45; Mismatches 105; Indels 14; Gaps 7;  
QY 1 MYLLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDE---LFSHLHPLLGDAMREIKGI 57  
| | : ||||| : | | : | | | : | | | : | | : | :  
Db 1 MLLVFDVGNNTNMVLGIYKGDKLVNRYRIKTDREKTSDEYGILISNLFYDYNVNISDIDDV 60  
| | : ||||| : | | : | | | : | | | : | | : | :  
QY 58 GVASVPTQNTVIERFSQKYFHISPIWV--KAKNGCVKWNV--NPSEVGADRVANVVAE 113  
: ||||| : | | | : | | : | | | : | | : | | : | |  
Db 61 IISVVPNVMHSLNFCIKYCKKQPLIVGPQIKTGL---NIKVDNPKQVGADRVNAVAG 117  
| | : ||||| : | | : | | | : | | : | | : | | : | |  
QY 114 VKEYGKNGIIDMGTTATVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKLPLVEVKPAD 172  
: || : ||||| : | | | : | | | : | | : | | : | | : | |  
Db 118 IEYGAPSLVDFGTATTFCAISEKGEYLGGTIAPGIKISSEALFQSASKLPRVELAKPG 177  
| | : ||||| : | | : | | | : | | : | | : | | : | |  
QY 173 FVVGKDTNIRLVNGSVVALEGIIGRIKEVYG--DLPVVLTTGGQSKIVKDMIKH-EI 229  
: | : || : | : | : | : | : | : | : | : | : | : | :  
Db 178 MTICKSTVSAMQSGIITYGYGLVDKIIISIMKKELNCDDVKVIATGGGLAKLIASETKSIDY 237  
| | : || : | : | : | : | : | : | : | : | : | : | :  
QY 230 FDEDLTIKGV 239  
| | : || : | : | : | : | : | : | : | : | : | : | :  
Db 238 VDGFLTLEGL 247

RESULT 14  
AAU91153  
ID AAU91153 standard; Protein; 258 AA.  
XX  
AC AAU91153;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Rhodobacter capsulatus pantothenate kinase Coax.  
XX  
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX  
OS Rhodobacter capsulatus.  
XX  
PN WO200216601-A2.  
XX  
PD 28-FEB-2002.  
XX  
PF 24-AUG-2001; 2001WO-US26531.  
XX  
PR 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX  
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX  
PI Yocum RR, Patterson TA;  
XX  
DR WPI; 2002-269358/31.  
XX  
PT Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -  
XX  
PS Claim 10; Page 71-72; 128pp; English.  
XX  
CC The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.  
XX  
SQ Sequence 258 AA;  
Query Match 23.1%; Score 295.5; DB 23; Length 258;  
Best Local Similarity 33.7%; Pred. No. 5.6e-24;  
Matches 86; Conservative 48; Mismatches 96; Indels 25; Gaps 10;  
QY 1 MYLLVDVGNTHSVFSITEDGKTF-RRWRLSTGVFQTEDELFSLHPLLGDAMREIKIG- 58  
| | : ||||| : | | : | | | : | | | : | | : | :  
Db 1 MLLCIDCGNTNVFSVW-DGTDFAATWRIATDHRRTADEYFVWLNTLM-----QLKGLQG 54  
| | : ||||| : | | : | | | : | | : | | : | | : | |  
QY 59 -----VASVPTQNTVIERFSQKYFHISPIWVKAKNGC---VKWNKPNSEVGADRVAN 109  
: | : || : | : | : | | : | | : | | : | | : | |  
Db 55 RISEAIIISTAPRVVFNLRVLCNRYFDCRP-YVVGKPGCELPVAPRVDPGTTVGPDLVN 113  
| | : ||||| : | | : | | | : | | : | | : | | : | |  
QY 110 VVAFVKEYGKNGIIDMGTTATVDLVV-NGSYEGGAILPGFFMMVHSLFRGTAKLPLVEV 168  
| | : ||||| : | | : | | | : | | : | | : | | : | |  
Db 114 TVAGYDRHGGDLIVVDFGTATTFDVVAPDGAYIGVGIAPGVNLSLEALHMAAALPHVDV 173  
| | : ||||| : | | : | | | : | | : | | : | | : | |  
QY 169 KPADFVVGKDTNIRLVNGSVVALEGIIGRIK-EVYGDLPVLTGGQSKIVKDMIKH 227  
: | : || : | : | : | : | : | : | : | : | : | : | :  
Db 174 TKPQGVIGTNTVACIQSGVYWGVIIGLVEGIVRQIRMERDRPMKVIATGGLASLFD--LGF 231  
| | : || : | : | : | : | : | : | : | : | : | : | :  
QY 228 EIFD---EDLTIKGV 239  
: | : || : | : | : | : | : | : | : | : | : | : | :



Db232DLFDKVEDDLTMHGL246

Db239HTVDHYDQHLTLOGL253

Search completed: June 24, 2003, 21:46:17  
Job time : 26.2563 secs

RESULT 15  
AAU91152  
IDAAU91152standard; Protein; 272 AA.  
XX  
ACAAU91152;  
XX  
DT05-JUN-2002 (first entry)  
XX  
DEMycobacterium tuberculosis pantothenate kinase Coax.  
XX  
KWPantothenate kinase; Coax; antibiotic; antimicrobial;  
KWpantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX  
OSMycobacterium tuberculosis.  
XX  
WO200216601-A2.  
J28-FEB-2002.  
XX  
PF24-AUG-2001; 2001WO-US26531.  
XX  
PR24-AUG-2000; 2000US-227860P.  
PR20-MAR-2001; 2001US-0813453.  
XX  
PA(OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX  
PIYocum RR, Patterson TA;  
XX  
DRWPI; 2002-269358/31.  
XX  
PTIdentifying potential antibiotic or antimicrobial agent, comprises  
PTcontacting composition comprising pantothenate kinase (Coax) protein  
PTwith test compound and identifying inhibitor of the Coax protein  
XX  
PSClaim 10; Page 70-71; 128pp; English.  
XX  
CCThe invention describes assays for identifying a (potential) antibiotic  
CCcomprising contacting an assay composition comprising a pantothenate  
CCkinase (Coax) protein with a test compound, and determining the ability  
CCof the test compound to inhibit the activity of the Coax protein, an  
CCessential enzyme for the production of coenzyme A. Coax protein is a  
CCvaluable target for identifying bactericidal compounds. Coax modulating  
CCagents can be used in an infectious animal model to determine the  
CCefficacy, toxicity, or side effects of treatment with such an agent. This  
CCis the amino acid sequence of a pantothenate kinase (Coax) protein  
described in the invention.  
XX  
SQSequence272AA;

Query Match22.3%; Score286; DB23; Length272;  
Best Local Similarity31.4%; Pred. No.6.7e-23;  
Matches80; Conservative49; Mismatches108; Indels18; Gaps7;

QY1MYLLVDVGNTHSVFSI---TEDGKTFRRWRRLSTGVFQTEDELFSLHPLGLGDAMREIKG56  
| | : | | | | | : | | : | | : | | | | : | | : | | : |  
Db1MLLAIDVRNTHTVVGLSGMKEHAKVVQQWRIRTESEVTADELALTIDGLIGEDSERLTG60

QY57IGVASVVPQTNTVIERFSQKYFHISP---IWVKAKGCVKWNKPNPSEVGADRVANVAVF113  
| | : | | : | | : | | : | | : | | : | | : | | : | | : |  
Db61TAALSTVPVSLHEVRIMLDQYWPSPVPHVLIIEPGVGTG-IPLLVDNPKEVGADRVNCLAA119

QY114VKEYGKNGIIMDGTATTVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKPLVEVKPAD172  
: | | : | | : | | : | | : | | : | | : | | : | | : | | : |  
Db120YDRFRKAAIVVDFGSSICVDVVSAAKGEFLGGAIPAGVQVSSDAAAARSAAALRRVELARPR179

QY173FWVGKDTENIRLGVVNGSVYALEGIIGRIKE-----VYGDLPVVLTTGGQSKIYKDMIK226  
| | | : | | : | | : | | : | | : | | : | | : | | : | | : |  
Db180SVVGKNTVECMQAGAVFGFAGLVLDGLVGRIREDEVSGFSVDHDAIVATGHTAPLPLPEL-238

QY227HEI--FDEDLTIKGV239

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:33:11 ; Search time 4.27862 Seconds  
(without alignments)  
2384.688 Million cell updates/sec

Title: US-09-813-453A-9  
Perfect score: 1281  
Sequence: 1 MYLLVDVGNTHSVFSITEDG.....HEIFDEDLTIKGVHFCFGD 246

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues  
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query %		Length	DB	ID	Description
	Score	Match				
1	369.5	28.8	233	1	YACB_BACSU	P37564 bacillus su
2	152	11.9	267	1	BAF_BORPE	Q45338 bordetella
3	94	7.3	397	1	AAT_STRVG	Q60013 streptomyce
4	94	7.3	1355	1	RPOD_ANASP	P22705 anabaena sp
5	93.5	7.3	717	1	RPOD_NOSCO	P14564 nostoc comm
6	92	7.2	330	1	Y840_METJA	Q58250 methanococc
7	91.5	7.1	429	1	AROA_METJA	Q57925 methanococc
8	90.5	7.1	989	1	RPOC_LEUME	P94892 leuconostoc
9	89.5	7.0	989	1	RPOC_LEUPS	P94899 leuconostoc
10	89	6.9	675	1	NUAM_ACACA	Q37373 acanthamoeb
11	88.5	6.9	489	1	LE11_ARCFU	O29305 archaeglob
12	88	6.9	545	1	CH60_AQUAE	O67943 aquifex ae
13	88	6.9	1206	1	RPOC_BACHD	Q929m1 bacillus ha
14	87	6.8	1004	1	RPOC_OENOE	P95405 oenococcus
15	86	6.7	386	1	CTBP_DROME	O46036 drosophila
16	86	6.7	1213	1	RPOC_STRPY	P95816 streptococc
17	85.5	6.7	532	1	FLID_BACHD	Q9k6w0 bacillus ha
18	85.5	6.7	608	1	HSCA_BUCAP	O51883 buchnera ap
19	85.5	6.7	880	1	CADE_XENLA	P33148 xenopus lae
20	85.5	6.7	1201	1	RPOC_LISMO	Q8ya96 listeria mo
21	85	6.6	545	1	CH60_PORGI	P77879 listeria in
22	84.5	6.6	1201	1	RPOC_LISIN	Q00184 escherichia
23	84	6.6	637	1	TRG5_ECOLI	P96177 weissella h
24	84	6.6	1046	1	RPOC_WEIHE	Q58142 methanococc
25	83.5	6.5	393	1	Y732_METJA	Q9rdv0 prochloroco
26	83.5	6.5	747	1	PSAB_PROMA	P77916 pyrococcus
27	83.5	6.5	771	1	DPOL_PYRAB	P15989 gallus gall
28	83.5	6.5	3137	1	CA36_CHICK	P19367 homo sapien
29	83	6.5	917	1	HXK1_HUMAN	P11568 acidaminoco
30	82.5	6.4	260	1	HGDC_ACIFE	P33655 clostridium
31	82.5	6.4	596	1	PRIM_CLOAB	P39987 saccharomyc
32	82.5	6.4	644	1	YED0_YEAST	P39002 saccharomyc
33	82.5	6.4	694	1	LCF3_YEAST	

34	82.5	6.4	1047	1	CARB_THEAC	Q9hk17 thermoplasm
35	82	6.4	437	1	SUCB_NEOFR	P53587 neocallimas
36	82	6.4	541	1	CH60_ANAPH	O34191 anaplasma p
37	82	6.4	600	1	DNAK_MYCHY	Q49539 mycoplasma
38	82	6.4	611	1	HSCA_BUCAI	P57660 buchnera ap
39	81.5	6.4	336	1	G3PC_PETCR	P26519 petroselinu
40	81.5	6.4	519	1	GALT_THETN	Q8r8r6 thermoanaer
41	81	6.3	1052	1	RPOC_BACAN	P77819 bacillus an
42	80.5	6.3	335	1	G3P_CHLMU	Q9pjn6 chlamydia m
43	80.5	6.3	517	1	LEUI_NEIMA	Q9juk6 neisseria m
44	80.5	6.3	591	1	CO8B_HUMAN	P07358 homo sapien
45	80.5	6.3	1055	1	RPOC_PEDAC	P77917 pediococcus

ALIGNMENTS

RESULT 1

YACB\_BACSU  
ID YACB\_BACSU STANDARD; PRT; 233 AA.

AC P37564;  
DT 01-OCR-1994 (Rel. 30, Created)  
DT 01-OCR-1994 (Rel. 30, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein yacB.

GN YACB.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;

RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX MEDLINE=96051385; PubMed=7584024;

RA Ogasawara N., Nakai S., Yoshikawa H.;

RT "Systematic sequencing of the 180 kilobase region of the Bacillus

subtilis chromosome containing the replication origin.";

RL DNA Res. 1:1-14(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX MEDLINE=98044033; PubMed=9384377;

RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,

RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,

RA Borriass R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,

RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,

RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,

RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,

RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,

RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,

RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,

RA Guiseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,

RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,

RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,

RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,

RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,

RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,

RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,

RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,

RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,

RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,

RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,

RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,

RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,

RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,

RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,

RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,

RA Viari A., Wambutt R., Wedler E., Wedler H., Weltzenegger T.,

RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,

RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;

RT "The complete genome sequence of the Gram-positive bacterium Bacillus

subtilis.";

RL Nature 390:249-256(1997).

-!- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.





```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D50624; BAA09299.1; -
CC HSP; Q56232; 1BJW.
CC InterPro; IPR001176; ACC_synthase.
CC InterPro; IPR004839; Aminotransf1/2.
CC InterPro; IPR004838; NHtransf_1.
CC Pfam; PF00155; aminotran_1.2; 1.
CC PRINTS; PR00753; ACCSYNTHASE.
CC PROSITE; PS00105; AA_TRANSFER_CLASS_1; 1.
CC Transferrase; Aminotransferase; Pyridoxal phosphate.
KW BINDING 236 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT BINDING 236
FT SEQUENCE 397 AA; 42381 MW; EEFEDCEB7D923065 CRC64;
-----
Query Match 7.3%; Score 94; DB 1; Length 397;
Best Local Similarity 24.5%; Pred. No. 1.3;
Matches 60; Conservative 38; Mismatches 101; Indels 46; Gaps 15;
QY 3 LLDVGN-THSVFSTEDGKTFRRWRLSTGVFQTEDELFSLH-----HPLLGDAMR 52
DB 167 LFVSPNPTGSVYS-EADAKAIGEWAAEHLWLVTDEIYEHLYGEAKFTSLPVLVPALR 225
QY 53 EIKGIGVASVPTQNTVIERFSQYFHSPIWVAKNGCVK--WNVKNPSEVGADRVANV 110
DB 226 D-KCIIV-----NGVAKTYAMTGRVG--WVIAPQDVKAATNLQSHATSNVSNVAQV 275
QY 111 VAFVKEYGKNGIIDMGTA-----TTVDLV--VNGSY-----EGG-AILPGFFMVHSLF 157
DB 276 AALAAVSGNLDAAVEMRKAFDRRRQTMVKMLNEIDGVFCPTPEGAFYAYPSVKELLGKEI 335
QY 158 RGTAKLPLVEVKPADFVVGKDTENIRLGVVNGSVYALEGIIGRIKEVYGDLPVVLTTGGQ 217
DB 336 RG--KRQSSVELAALIL-----DEVEAVVPGEAFGTPGYL-RLSYALGDEDLV--EGV 385
QY 218 SKIVK 222
DB 386 SRIQK 390
-----
RESULT 4
RPOD_ANASP STANDARD; PRT; 1355 AA.
AC P22705;
DT 01-AUG-1991 (Rel. 19, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA-directed RNA polymerase delta chain (EC 2.7.7.6).
GN RPOC2 OR ALR1596.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
RN [2]
RP SEQUENCE OF 1-63 FROM N.A.
RX MEDLINE=91258327; PubMed=1904436;
RA Bergsland K.J., Haselkorn R.;
RT "Evolutionary relationships among eubacteria, cyanobacteria, and
```

RL J. Bacteriol. 171:1967-1973(1989).  
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
CC SUBSTRATES.  
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
CC {RNA}(N).  
CC -!- SUBUNIT: IN CYANOBACTERIA THE RNA POLYMERASE IS COMPOSED OF FOUR  
CC SUBUNITS: ALPHA, BETA, GAMMA, AND DELTA.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M29747; AAA25518.1; -.  
CC PIR; B32838; B32838.  
CC DR HSSP; Q9KWU6; 1HQM.  
CC -!- InterPro; IPR000722; RNA\_pol\_A.  
CC -!- InterPro; IPR002879; RNA\_pol\_A2.  
CC -!- Pfam; PF00623; RNA\_pol\_A; 1.  
CC -!- Pfam; PF01854; RNA\_pol\_A2; 1.  
CC KW Transferase; Transcription; DNA-directed RNA polymerase.  
CC FT NON\_TER 717 717  
CC SQ SEQUENCE 717 AA; 78244 MW; E5505A9FA37D3C91 CRC64;

Query Match 7.3%; Score 93.5; DB 1; Length 717;  
Best Local Similarity 22.2%; Pred. No. 2.9;  
Matches 61; Conservative 35; Mismatches 94; Indels 85; Gaps 13;  
QY 4 LVDVGNTH-----SVESITDGKTF--RRWRLSTGVF-----QTEDELFSHL 43  
Db 205 LVDVSQVYSGFDCGTPSLRPMTEGAKTLIPLATRLMGRVIGEDVLPVTKVEIAARN 264  
QY 44 HPLLGDAMREIKGIGVASVPTQNTVIERFSQKYFHISPIWVKAKNG----CVKWNVKNP 99  
Db 265 SPISEDLAKKIEKSGVGEVVR-----SPLTCEAARSVCQHCHYGS LAHA 309  
QY 100 SEVGADRVANVAVFKEYGKNGIIDMGATTVDLVVNGSYEGGAILPGFFMMVHSLFRG 159  
Db 310 SMVDLGEAVGIIA-AQSIGEPGTQLTMRFTHT-----GGVFTGEVAQQVRSKIDG 358  
QY 160 TAKLP-----LVEVKPADVGVGKDTTEENIRLGVVNGS-VYAL 195  
Db 359 TVKLPRKLKTRTYRTRHGEDALYVEANGIMLEPTK--VGDVTPENQEVHLTQGSTLYVF 416  
QY 196 EGIIGRIKEVYGLPVLTGGQS-----KIVKDM 224  
417 DG--NKVNKQV-LLAEEVALGGRTTRTNTTEKAVKDV 448

RESULT 6  
Y840\_METJA  
ID Y840\_METJA STANDARD; PRT; 330 AA.  
AC Q58250;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein MJ0840.  
GN MJ0840.  
OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
OC Methanocaldococcales; Methanocaldococcus.  
OX NCBI\_TaxID=2190;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
RX MEDLINE=96337999; PubMed=868087;  
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,  
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,

RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
RT jannaschii.";  
RL Science 273:1058-1073(1996).  
CC -!- SIMILARITY: STRONG, TO M.VANNIELII HYPOTHETICAL PROTEIN IN ARGG  
CC 5'REGION.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U67528; AAB98845.1; -.  
CC TIGR; MJ0840; -.  
CC InterPro; IPR002756; DUF66.  
CC Pfam; PF01897; DUF66; 1.  
CC ProDom; PD024636; DUF66; 1.  
CC KW Hypothetical protein; Complete proteome.  
CC FT SEQUENCE 330 AA; 37376 MW; 5A3EAA6D2125D829 CRC64;

Query Match 7.2%; Score 92; DB 1; Length 330;  
Best Local Similarity 20.4%; Pred. No. 1.5;  
Matches 66; Conservative 51; Mismatches 106; Indels 100; Gaps 14;  
QY 1 MYLLVDVGNTHSVFSITE-DGKTFRRWRLSTGVFQTEDELFSHLPLGDAMREIKGIGV 59  
Db 4 MILGIDIGGANT--KITEIEGDNYKIHIIYFPMWKKKDEL-EDLLKNYNDNVYVALVMT 60  
QY 60 ASVPTQNT-----VIERFSQKYPHISPIWVKAKNG---CVKWNVKNPSEVGADRVAN 109  
Db 61 AELADCYKTKKEGVEDIIDKV-ERAFN-CPVYVFDVNGNFLTSEEAKKNYLDVSASNWA 118  
QY 110 VVAFVKEYGKNG-IIDMGATT-----VDLVVNGS--YEGGAILPGFF 150  
Db 119 TAKFVAEFIKDSCILVDMGSTTDDIPIKDEVLAEKTDLDRLMNNQLVYVGTLPVSF 178  
QY 151 MVHSLFRG-----TAKLPVLEVVKPADVGVGKDTTEENIRLGVVNGSVYALEGI 198  
Db 179 LANKIEFRGKLTNLSSEYFAITADISL-----ILNKITEEDYTCDDPDGAGKDFESC 230  
QY 199 IGRIKEV-----YGDLPVLTGGQSKI 220  
Db 231 LTRLRVLCADREVMKDELIDFANKLYNKLELIRENVDTIAKRYNLDNVITGLGEEI 290  
QY 221 VKDMIKH-----EIFDEDLTI 236  
Db 291 LKDALDEYNIISIKETYGKDVSL 313

RESULT 7  
AROA\_METJA  
ID AROA\_METJA STANDARD; PRT; 429 AA.  
AC Q57925;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Probable 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19)  
DE (5-enolpyruvylshikimate-3-phosphate synthase) (EPSP synthase) (EPSPS).  
GN AROA OR MJ0502.  
OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
OC Methanocaldococcales; Methanocaldococcus.  
OX NCBI\_TaxID=2190;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;





DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE DNA-directed RNA polymerase beta' chain (EC 2.7.7.6) (Transcriptase  
DE beta' chain) (RNA polymerase beta' subunit) (Fragment).  
GN RPOC.  
OS Leuconostoc pseudomesenteroides.  
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
OX NCBI\_TaxID=33968;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCDO 768;  
RX MEDLINE=97016803; PubMed=8863429;  
RA Morse R., Collins M.D., O'Hanlon K., Wallbanks S., Richardson P.T.;  
RT "Analysis of the beta' subunit of DNA-dependent RNA polymerase does  
RT not support the hypothesis inferred from 16S rRNA analysis that  
RT Oenococcus oeni (formerly Leuconostoc oenos) is a tachytelic  
RT (fast-evolving) bacterium.";  
RL Int. J. Syst. Bacteriol. 46:1004-1009(1996).  
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
CC SUBSTRATES.  
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +  
CC [RNA](N).  
CC -!- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE  
CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1  
CC BETA' CHAIN.  
CC -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X95812; CAA65079.1; .  
DR HSSP; Q9KWU6; 1HQM.  
DR InterPro; IPR000722; RNA\_pol\_A.  
DR InterPro; IPR002879; RNA\_pol\_A2.  
DR Pfam; PF00623; RNA\_pol\_A; 1.  
DR Pfam; PF01854; RNA\_pol\_A2; 1.  
KW Transferase; DNA-directed RNA polymerase; Transcription.  
FT NON\_TER 1  
FT NON\_TER 989 989  
FT SEQUENCE 989 AA; 109914 MW; 55B1CD8CA8739D06 CRC64;  
SQ  
Query Match 7.0%; Score 89.5; DB 1; Length 989;  
Best Local Similarity 21.0%; Pred. No. 9.3;  
Matches 58; Conservative 41; Mismatches 106; Indels 71; Gaps 12;  
6 DVGNTSVFSITEDGKTFRRWRLSTGVFQTEDELFSLHPLLDGDMREIKGIGVASVVP 65  
579 DLGYEKS-----TESGLT-----VAMTDVTDLKERPAILEDHQNVA----- 615  
66 QNTVIERF-----SQYFHISPIWVAKN-----GCVKWNVKNPSEVGADRV--NVV 111  
616 --TVTKQFRRGLITDDERYQVRVEIWTAKDIIQDKLIESFEPTNPIFMMQDSGARGNIS 673  
112 AFVKEYGKNGIIDMGTAFTVDLVVNGSYEGGAILPGFMMVHSLFRGTAKLP----- 164  
674 NFVQLAGMRGLMAGPG-GKIIEPLVTANFREGLTVMEMPISTHGARKGMSDTALKTANSV 732  
165 -----LVEVKPADFVV---GKDTEENIRLGVVNGSVYALEGIIIGRIKEVYGLPVLTTGG 216  
733 YLTRLVDV-AQDVIVREFDNDSDRGVAVKAIMDGTSVVEPLYDRILGRYAMKSVFDEPT 791  
217 QSKIVKDMIKHEIFDED-----LTIKGVY 240  
792 GEKIVS---RNEMIDEDVAKAIVNAGIEEVTIRSVF 824

RESULT 10

NUAM\_ACACA

ID NUAM\_ACACA STANDARD; PRT; 675 AA.  
AC Q37373;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE NADH dehydrogenase, subunit 11 (EC 1.6.5.3) (NADH dehydrogenase  
DE (Ubiquinone)) (Ubiquinone reductase) (Type I dehydrogenase) (Complex I  
DE dehydrogenase).  
GN NAD11.  
OS Acanthamoeba castellanii (Amoeba).  
OG Mitochondrion.  
OC Eukaryota; Acanthamoebidae; Acanthamoeba.  
OX NCBI\_TaxID=5755;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 30010 / NEFF;  
RX MEDLINE=95147275; PubMed=7844823;  
RA Burger G., Plante I., Loneragan K.M., Gray M.W.;  
RT "The mitochondrial DNA of the amoeboid protozoon, Acanthamoeba  
RT castellanii: complete sequence, gene content and genome  
RT organization.";  
RL J. Mol. Biol. 245:522-537(1995).  
CC -!- FUNCTION: THIS IS THE LARGEST SUBUNIT OF COMPLEX I AND IT IS A  
CC COMPONENT OF THE IRON-SULFUR (IP) FRAGMENT OF THE ENZYME. IT MAY  
CC FORM PART OF THE ACTIVE SITE CREVICE WHERE NADH IS OXIDIZED (BY  
CC SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.  
CC -!- COFACTOR: BINDS 1 2FE-2S CLUSTER AND 1 4FE-4S CLUSTER (Potential).  
CC -!- SUBUNIT: COMPLEX I IS COMPOSED OF ABOUT 30 DIFFERENT SUBUNITS (BY  
CC SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: MATRIX AND CYTOPLASMIC SIDE OF THE  
CC MITOCHONDRIAL INNER MEMBRANE (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE COMPLEX I 75 kDa SUBUNIT FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U12386; AAD11824.1; .  
DR InterPro; IPR000283; Complex1\_75K.  
DR InterPro; IPR001041; Ferredoxin.  
DR Pfam; PF00111; fer2; 1.  
DR PROSITE; PS00641; COMPLEX1\_75K\_1; FALSE\_NEG.  
DR PROSITE; PS00642; COMPLEX1\_75K\_2; 1.  
DR PROSITE; PS00643; COMPLEX1\_75K\_3; 1.  
KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Iron-sulfur; 4Fe-4S.  
FT METAL 25 25 IRON-SULFUR (4FE-4S) (POTENTIAL).  
FT METAL 36 36 IRON-SULFUR (4FE-4S) (POTENTIAL).  
FT METAL 47 47 IRON-SULFUR (4FE-4S) (POTENTIAL).  
FT METAL 50 50 IRON-SULFUR (4FE-4S) (POTENTIAL).  
FT METAL 64 64 IRON-SULFUR (4FE-4S) (POTENTIAL).  
FT METAL 100 100 IRON-SULFUR (4FE-4S) (POTENTIAL).  
FT METAL 103 103 IRON-SULFUR (4FE-4S) (POTENTIAL).  
FT METAL 109 109 IRON-SULFUR (4FE-4S) (POTENTIAL).  
FT METAL 148 148 IRON-SULFUR (4FE-4S) (POTENTIAL).  
FT METAL 151 151 IRON-SULFUR (4FE-4S) (POTENTIAL).  
FT METAL 154 154 IRON-SULFUR (4FE-4S) (POTENTIAL).  
FT METAL 198 198 IRON-SULFUR (4FE-4S) (POTENTIAL).  
SQ SEQUENCE 675 AA; 77554 MW; 3870F3BC9E8B9807 CRC64;  
Query Match 6.9%; Score 89; DB 1; Length 675;  
Best Local Similarity 19.1%; Pred. No. 6.5;  
Matches 54; Conservative 37; Mismatches 93; Indels 98; Gaps 11;

QY 22 TFRWRRL-STGVFQTEDELFSLHPLLDGDMREIKGIGVASVVPQTNTVIE-----R 72

Db 211 TSRPWELKSYNSIDVLDLSHSNIRV-----DIRGTMKIMRILPRVNSELDNEDWITDKIR 263

QY 73 FSQKYFHISPIW---VKAKNGCVKWNVKNPSEVGADRVANVAVFVKEY----- 117



Db 264 FSYDSFRQRRLYDPMVKISGSLKIGWK-----KAMLFIKKFFCNFLGFNHSF 312  
QY 118 ----GKNGIIDIIMGTATVD--LVVNGSYEGGAILPGFMMVHSLFRGTA---KLPLVE 167  
Db 313 IPLRGYIGDYLDLEITYTFKFKLLNGS-----NFFLPSSSYNDLTALYSFNTPLTR 364  
QY 168 VKPADFVVVGKDTNEINIRLGVVGSV-----192  
Db 365 LDEGDFCILLDVNLRVELPIVNSRIKQLVSKKMLPVFLGFSYFNENYFVKHISNSSKTL 424  
QY 193 YALEG---IIGRIKEVYGDLPVLTGGQSKIIVKDMIKHEIFD 231  
Db 425 HVLEGSHWLSAKISKFSKPIFLIGDSSLLKGSLLVPLEN 466

RESULT 11  
LE11\_ARCFU STANDARD; PRT; 489 AA.  
O29305;  
15-JUN-2002 (Rel. 41, Created)  
15-JUN-2002 (Rel. 41, Last sequence update)  
15-JUN-2002 (Rel. 41, Last annotation update)  
2-isopropylmalate synthase 1 (EC 4.1.3.12) (Alpha-isopropylmalate synthase 1) (Alpha-IPM synthetase 1).  
LEUA1 OR AF0957.  
OS Archaeoglobus fulgidus.  
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;  
OC Archaeoglobaceae; Archaeoglobus.  
OX NCBI\_TaxID=2234;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
RX MEDLINE=98049343; PubMed=9389475;  
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E., Klenk K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D., Ketchum K.A., Kerlavage A.R., Graham D.E., Kyrpides N.C., Richardson D.L., Quackenbush J., Lee N.H., Sutton G.G., Gill S., Fleischmann R.D., McQuinn K., Adams M.D., Loftus B., Kirkness E.F., Dougherty B.A., McKenney K., Badger J.H., Glodek A., Zhou L., Peterson S., Reich C.I., McNeil L.K., Weidman J.F., McDonald L., Utterback T., Overbeek R., Gocayne J.D., Weidman J.F., Kaine B.P., Sykes S.M., Cotton M.D., Spriggs T., Artach P., Bowland C., Garland S.A., Sadow P.W., D'Andrea K.P., Fraser C.M., Smith H.O., Woese C.R., Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
RT "The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon Archaeoglobus fulgidus.";  
RL Nature 390:364-370(1997).  
CC -!- FUNCTION: Catalyzes the condensation of the acetyl group of acetyl-CoA with 3-methyl-2-oxobutanoate (2-oxoisovalerate) to form 3-carboxy-3-hydroxy-4-methylpentanoate (2-isopropylmalate).  
CC -!- CATALYTIC ACTIVITY: 3-carboxy-3-hydroxy-4-methylpentanoate + CoA -> acetyl-CoA + 3-methyl-2-oxobutanoate + H(2)O.  
CC -!- PATHWAY: Leucine biosynthesis; first step.  
CC -!- SIMILARITY: BELONGS TO THE ALPHA-IPM SYNTHETASE / HOMOCITRATE SYNTHASE FAMILY. LEUA 1 SUBFAMILY.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AE001038; AAB90286.1; --  
TIGR; AF0957; --  
DR InterPro; IPR002034; AIPM/Hcit\_synth.  
DR InterPro; IPR000891; HMGL-like.  
DR Pfam; PF00682; HMGL-like; 1.  
DR PROSITE; PS00815; AIPM\_HOMOCIT\_SYNTH\_1; 1.  
DR PROSITE; PS00816; AIPM\_HOMOCIT\_SYNTH\_2; 1.  
KW Leucine biosynthesis; Lyase; Complete proteome.

SQ SEQUENCE 489 AA; 53345 MW; 9E84B204DE796F84 CRC64;  
Query Match 6.9%; Score 88.5; DB 1; Length 489;  
Best Local Similarity 22.5%; Pred. No. 4.8;  
Matches 48; Conservative 37; Mismatches 91; Indels 37; Gaps 7;  
QY 45 PLLGD-AMREIKGIGVASVPTQNTVIERFSQKYFHISPIWVKAKGCVKWNKNPSEVG 103  
Db 274 PIVGDNATFTHESGI-----HTSALFRDAKSYEPISPEVVGKRVIVLGLKHAGRASVE 325  
QY 104 A-----DRVANVAVFVKEYGKNG-IIDMGATTTVDLVVNGSYEGGAILPGFFM 151  
Db 326 AIMNELGYKATPEQMKEILARIKEIGDKGKRVTDADVRTIETVLQIKREKKVKLEDL-- 383  
QY 152 MVHSLFRGTAFLPLVEVKPADFVVGKDTNEINIRLGVVNGSVYALEGIIGRIKEVYGDLPV 211  
Db 384 ---AIFSGKNVMPMASVK-----LKIDGQERIEAAVGLGPVDAAINAIRRAIKEFADIKL 435  
QY 212 V-----LTGQSKIIVKDMIKHEIFDEDLTIKG 238  
Db 436 VSYHVDATGTDALVDVVVQLKDKNKIVTARG 468

RESULT 12  
CH60\_AQUAE STANDARD; PRT; 545 AA.  
ID CH60\_AQUAE  
AC O67943;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein).  
GN GROL OR MOPA OR GROEL OR AQ\_2200.  
OS Aquifex aeolicus.  
OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;  
OC Aquifex.  
OX NCBI\_TaxID=63363;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VF5; PubMed=9537320;  
RX MEDLINE=98196666; PubMed=9537320;  
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L., Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R., Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
RA "The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.";  
RT Nature 392:353-358(1998).  
RL -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS CONDITIONS (BY SIMILARITY).  
CC -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF 7 SUBUNITS (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AE000777; AAC07897.1; --  
DR HSSP; P06139; 1GRL.  
DR InterPro; IPR001844; Chaperonin\_Cpn60.  
DR InterPro; IPR002423; Cpn60/TCP-1.  
DR Pfam; PF00118; cpn60\_TCP1; 1.  
DR PRINTS; PR00298; CHAPERONIN60.  
DR PRINTS; PR00304; TCOMPLEXTCP1.  
DR PROSITE; PS00296; CHAPERONINS\_CPN60; 1.  
KW Chaperone; ATP-binding; Complete proteome.  
SQ SEQUENCE 545 AA; 58794 MW; 6CB190ED6E8BC8B9 CRC64;  
Query Match 6.9%; Score 88; DB 1; Length 545;



modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X96384; CAA65248.1; --

HSSP; Q9KWU6; 1HQW.

InterPro; IPR000722; RNA\_pol\_A.

InterPro; IPR002879; RNA\_pol\_A2.

Pfam; PF00623; RNA\_pol\_A; 1.

Pfam; PF01854; RNA\_pol\_A2; 1.

Transferase; DNA-directed RNA polymerase; Transcription.

NON\_TER 1

NON\_TER 1004 1004

SEQUENCE 1004 AA; 111965 MW; 73750DF47F3A2C36 CRC64;

Query Match 6.8%; Score 87; DB 1; Length 1004;

Best Local Similarity 19.8%; Pred. No. 15;

Matches 48; Conservative 47; Mismatches 88; Indels 60; Gaps 11;

21 KTFRRRLSTGVFQTEDELFSHLHPLLDGAMREIKGIGVASVPTQNTVIERFSQKYFHI 80

624 KQFRGRL-----TDERYIQVTQTSQAQDKIKSMLIAS-----FDSK---- 662

81 SPIWVKAKNCVKWNPSEVGADRVANVAVFVKEYGKNGIIDMGTATVVDLVVNGSY 140

663 NPIFMMSDSG-ARGNISNSFOLAGMR--GLMA-----APNGKVI-----ELPVTANF 706

141 EGGAILPGFMMVHSLFRGTAKLP-----LVEVKPADVFVVGKD--TEENIRLG 186

707 REGLCVLEMFISTHGARKGMTDTALKTANSGLYTRRLVDVAQEVIVREDCGTDRGLDVS 766

187 VVNGSVYALEGIGRIKEVYGDLPV-----LTGGQSKIVKDMIKHEIFD---EDLTIK 237

767 AIMDGNIEVIEPLYDRILGRYAMKPVIDPKTGEVIAKKNQIMIDEHVADQIIDAGIQTVTIR 826

238 GYV 240

827 SIF 829

RESULT 15

CTBP\_DROME

ID CTBP\_DROME STANDARD; PRT; 386 AA.

AC O46036; O61283; Q9VG02;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

E C-terminal binding protein (CtBP protein) (dCtBP).

IN CTBP OR CG7583.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;

OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

SEQUENCE FROM N.A., FUNCTION, AND ALTERNATIVE SPLICING.

TISSUE=Embryo;

MEDLINE=98190087; PubMed=9524128;

RA Poortinga G., Watanabe M., Parkhurst S.M.;

RT "Drosophila CtBP: a Hairy-interacting protein required for embryonic

segmentation and hairy-mediated transcriptional repression.";

RL EMBO J. 17:2067-2078(1998).

RN [2]

SEQUENCE FROM N.A., AND FUNCTION.

RX MEDLINE=98192810; PubMed=9525852;

RA Nibu Y., Zhang H., Levine M.;

RT "Interaction of short-range repressors with Drosophila CtBP in the

embryo.";

RL Science 280:101-104(1998).

RN [3]

SEQUENCE FROM N.A.

RC STRAIN-Berkeley;

RX MEDLINE=20196006; PubMed=10731132;

Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
Fowler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
Glodek A., Gong F., Gorrell J.H., Gu Z., Hernandez J.R., Houck J.,  
Harris N.L., Harvey D., Heiman T.J., Howland T.J., Wei M.-H., Ibegwam C.,  
Hostin D., Houston K.A., Kachur K., Ke Z., Kennison J.A., Ketchum K.A.,  
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
"The genome sequence of Drosophila melanogaster.";  
Science 287:2185-2195(2000).  
-!- FUNCTION: CO-REPRESSOR TARGETING DIVERSE TRANSCRIPTION REGULATORS.  
HAIRY-INTERACTING PROTEIN REQUIRED FOR EMBRYONIC SEGMENTATION AND  
HAIRY-MEDIATED TRANSCRIPTIONAL REPRESSION.  
-!- SUBUNIT: CAN FORM HOMODIMERS. INTERACTS WITH HAIRY, KNIRPS, SNAIL,  
AND ENHANCER OF SPLIT M-DELTA. MAY BE INVOLVED IN TRANSCRIPTIONAL  
REPRESSION. INTERACTS ALSO WITH ADENOVIRUS E1A PROTEIN.  
-!- SUBCELLULAR LOCATION: Nuclear.  
-!- ALTERNATIVE PRODUCTS: A number of isoforms are produced by  
alternative splicing.  
-!- DEVELOPMENTAL STAGE: TRANSCRIPTS LEVELS INCREASE BOTH EARLY DURING  
OOGENESIS AND EMBRYOGENESIS, AND LATER IN PRE-PUPAE STAGES.  
TRANSCRIPTS ARE DETECTED IN THE GERMANIUM AND IN THE NURSE CELLS  
THROUGHOUT THE EARLY OGENIC STAGES, PEAKING IN STAGE 10 NURSE  
CELLS. THEY ARE LATER DUMPED INTO THE OOCYTE WHERE THEY ARE  
DISTRIBUTED UBIQUITOUSLY. THE MATERNAL TRANSCRIPT PERSIST  
THROUGHOUT THE EARLY EMBRYONIC STAGES. TRANSCRIPTS LEVELS DROP  
DURING GASTRULATION AND ARE MAINTAINED AT THIS LOWER LEVEL  
THROUGHOUT THE REMAINDER OF EMBRYOGENESIS.  
-!- SIMILARITY: BELONGS TO THE D-ISOMER SPECIFIC 2-HYDROXYACID  
DEHYDROGENASES FAMILY. BUT HIGHLY DIVERGENT.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AJ224690; CAA12074.1; --  
EMBL; AB011840; BAA25287.1; --  
EMBL; AE003698; AAF54891.1; --  
HSSP; P17584; 1DXY.  
FlyBase; FBgn0020496; CtBP.  
InterPro; IPR002162; D\_2hydac\_dh.

```
DR Pfam; PF00389; 2-Hacid_DH; 1.
DR Pfam; PF02826; 2-Hacid_DH_C; 1.
DR PROSITE; PS00065; D_2_HYDROXYACID_DH_1; FALSE_NEG.
DR PROSITE; PS00670; D_2_HYDROXYACID_DH_2; FALSE_NEG.
DR PROSITE; PS00671; D_2_HYDROXYACID_DH_3; FALSE_NEG.
KW Alternative splicing; Nuclear protein.
FT DOMAIN 362 365 POLY-ALA.
FT CONFLICT 299 301 MISSING (IN REF. 2).
FT CONFLICT 371 371 A -> S (IN REF. 1).
SQ SEQUENCE 386 AA; 42251 MW; 89B9527BFF5224EC CRC64;

Query Match      6.7%; Score 86; DB 1; Length 386;
Best Local Similarity 20.5%; Pred. No. 5.9;
Matches 53; Conservative 41; Mismatches 116; Indels 48; Gaps 8;

QY 11 HVSFSTEDGKTFRRWRLSTGVFQTEDELFSLHPLLGDAMREIKGIGVASVVPQTNTVI 70
Db 11 HVSFSTEDGKTFRRWRLSTGVFQTEDELFSLHPLLGDAMREIKGIGVASVVPQTNTVI 70
QY 77 HTILTKEDELEKFKALRIIVRGSGTDNIDVKAAGELGIAVCNVPGYGVEEADTTMCLI 136
Db 77 HTILTKEDELEKFKALRIIVRGSGTDNIDVKAAGELGIAVCNVPGYGVEEADTTMCLI 136
QY 71 ERFQKYFHS-----PIWVK-AKGCVKWVKNPSEVGADRVANVVAFA-VKEY 117
Db 71 ERFQKYFHS-----PIWVK-AKGCVKWVKNPSEVGADRVANVVAFA-VKEY 117
QY 137 LNLVRRTYWLANMVREGKKFTGPEQVREAAHGCARIRGDTLGLVGLGRIGSAVALRAKAF 196
Db 137 LNLVRRTYWLANMVREGKKFTGPEQVREAAHGCARIRGDTLGLVGLGRIGSAVALRAKAF 196
QY 118 GKNGIITDMGTATTVDLVVNGSYEGGAILPGFFMMVHSLFRGTA----- 161
Db 118 GKNGIITDMGTATTVDLVVNGSYEGGAILPGFFMMVHSLFRGTA----- 161
QY 197 GFNVIFYDPYLPDGIDKSLG-----LTRYTLQDLLFQSDCVSLHCTLNEHNNHLLIN 248
Db 197 GFNVIFYDPYLPDGIDKSLG-----LTRYTLQDLLFQSDCVSLHCTLNEHNNHLLIN 248
QY 162 KLPLVEVKPADFVVGKDTTEENIRLGWVNGSVYALEGIIGRIK----EVYGDLPVVLGGQ 217
Db 162 KLPLVEVKPADFVVGKDTTEENIRLGWVNGSVYALEGIIGRIK----EVYGDLPVVLGGQ 217
QY 249 EFTIKQMRPGAFV-----NTARGGLVDDETALALAKQGRIRAAALDVHENEPYNVFOGA 303
Db 249 EFTIKQMRPGAFV-----NTARGGLVDDETALALAKQGRIRAAALDVHENEPYNVFOGA 303
QY 218 SKIVKDMI--KHEIFDED 233
Db 218 SKIVKDMI--KHEIFDED 233
QY 304 LKDAPNLICTPHAAFFSD 321
Db 304 LKDAPNLICTPHAAFFSD 321
```

Search completed: June 24, 2003, 21:51:15  
Job time : 5.27862 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:35:06 ; Search time 20.0638 Seconds  
(without alignments)  
2526.317 Million cell updates/sec

Title: US-09-813-453A-9  
Perfect score: 1281  
Sequence: 1 MYLLVDVGNTHSVFSITEDG.....HEIFDEDLTIKGVYHFCFGD 246

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

--Searched: 671580 seqs, 206047115 residues 671580  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_21.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1281	100.0	246	16 Q9WZY5	Q9wzy5 thermotoga
2	389.5	30.4	254	16 Q9KGH5	Q9kgh5 bacillus ha
3	378.5	29.5	259	16 Q8YAC5	Q8yac5 listeria mo
4	377.5	29.5	259	16 Q92F54	Q92f54 listeria in
5	371.5	29.0	258	2 Q9F985	Q9f985 bacillus st
6	360.5	28.1	273	16 Q97EB4	Q97eb4 clostridium
7	359.5	28.1	261	16 Q9A621	Q9a621 caulobacter
8	342.5	26.7	259	16 Q8XHL5	Q8xhl5 clostridium
9	342.5	26.7	265	16 Q9X8N6	Q9x8n6 streptomyc
10	324.5	25.3	255	16 Q8R7M2	Q8r7m2 thermoanaer
11	317.5	24.8	256	16 Q8RFE4	Q8rfe4 fusobacteri
12	289	22.6	274	16 Q9CD56	Q9cd56 mycobacteri
13	286	22.3	272	16 Q06282	Q06282 mycobacteri
14	285	22.2	273	16 Q83446	Q83446 treponema p
15	260	20.3	262	16 Q9RX54	Q9rx54 deinococcus
16	221	17.3	212	2 Q32514	Q32514 desulfovibr

17	204.5	16.0	224	16 Q98Q93	Q98q93 mycoplasma
18	203	15.8	262	16 Q51477	Q51477 borrelia bu
19	193.5	15.1	229	16 Q67753	Q67753 aquifex ao
20	179	14.0	223	16 Q92KY6	Q92ky6 helicobacte
21	165	12.9	223	16 Q25533	Q25533 helicobacte
22	150.5	11.7	592	16 Q9JW17	Q9jw17 neisseria m
23	149.5	11.7	295	16 Q8Y2M4	Q8y2m4 ralstonia s
24	146	11.4	592	16 Q9JXF1	Q9jxf1 neisseria m
25	145	11.3	248	16 Q9HWC1	Q9hwc1 pseudomonas
26	139.5	10.9	209	16 Q9PIA9	Q9pia9 campylobact
27	138	10.8	257	16 P74045	P74045 synechocyst
28	124.5	9.7	276	16 Q8YQD7	Q8yqd7 anabaena sp
29	103	8.0	242	16 Q9PC14	Q9pc14 xylella fas
30	101	7.9	322	2 Q8VVB8	Q8vvb8 streptococc
31	100.5	7.8	313	16 Q97FW3	Q97fw3 clostridium
32	100	7.8	831	16 Q9KZP4	Q9kzp4 streptomyc
33	98	7.7	315	16 Q51771	Q51771 borrelia bu
34	98	7.7	327	2 Q9RE12	Q9re12 lactobacill
35	98	7.7	431	16 Q8XNA5	Q8xna5 clostridium
36	96.5	7.5	816	16 Q8RB00	Q8rb00 thermoanaer
37	95	7.4	713	10 Q9FHG5	Q9fhg5 arabidopsis
38	95	7.4	780	16 Q9K917	Q9k917 bacillus ha
39	94.5	7.4	287	16 Q9KR61	Q9kr61 vibrio chol
40	94	7.3	1335	2 Q9LA54	Q9la54 escherichia
41	93	7.3	1335	2 Q9LA58	Q9la58 escherichia
42	92.5	7.2	320	16 Q97RW8	Q97rw8 streptococc
43	92	7.2	222	2 Q93PC4	Q93pc4 microscilla
44	92	7.2	298	16 Q92LP9	Q92lp9 rhizobium m
45	92	7.2	552	16 Q98AX9	Q98ax9 rhizobium l

ALIGNMENTS

RESULT 1

Q9WZY5 PRELIMINARY; PRT; 246 AA.  
ID Q9WZY5  
AC Q9WZY5;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Hypothetical protein TM0883.  
GN TM0883.  
OS Thermotoga maritima.  
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.  
OX NCBI\_TaxID=2336;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MSB8 / DSM 3109;  
RX MEDLINE=99287316; PubMed=10360571;  
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,  
Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,  
Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;  
RT "Evidence for lateral gene transfer between Archaea and Bacteria from  
genome sequence of Thermotoga maritima.";  
RL Nature 399:323-329(1999).  
DR EMBL; AE001754; AAD35964.1; -  
DR TIGR; TM0883; -  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMs; TIGR00671; baf; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 246 AA; 27154 MW; 9E0309AD462CF266 CRC64;

Query Match 100.0%; Score 1281; DB 16; Length 246;  
Best Local Similarity 100.0%; Pred. No. 2.2e-102;  
Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MYLLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDELFSLHPLGLDAMREIKGIGVA 60  
|||||

Db 1 MYLLVDVGNTHSVFSITEDGKTFRRWRRLSTGVFQTEDELFSHLHPLLGDMREIKGIGVA 60  
QY 61 SVVPTQNTVIERFSQKYFHISPIWVKAKNGCVKWNKNPSEVGADRVANVAVFVKEYGKN 120  
Db 61 SVVPTQNTVIERFSQKYFHISPIWVKAKNGCVKWNKNPSEVGADRVANVAVFVKEYGKN 120  
QY 121 GIIIDMGATTVDLVNNGSYEGGAILPGFFMMVHSLFRGTAKLPLVEVKPADFVVGKDTE 180  
Db 121 GIIIDMGATTVDLVNNGSYEGGAILPGFFMMVHSLFRGTAKLPLVEVKPADFVVGKDTE 180  
QY 181 ENIRLGVNNGSVYALEGIIGRIKEYGDLPPVLTGGQSKIIVKDMIKHEIFDEDLTIKGVY 240  
Db 181 ENIRLGVNNGSVYALEGIIGRIKEYGDLPPVLTGGQSKIIVKDMIKHEIFDEDLTIKGVY 240  
QY 241 HFCFGD 246  
Db 241 HFCFGD 246

RESULT 2  
9KGH5  
J Q9KGH5 PRELIMINARY; PRT; 254 AA.  
AC Q9KGH5;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical protein BH0086.  
GN BH0086.  
OS Bacillus halodurans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Bacillaceae; Bacillus.  
OX NCBI\_TaxID=86665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C-125 / JCM 9153;  
RX MEDLINE=20512582; PubMed=11058132;  
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,  
RA Horikoshi K.;  
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
halodurans and genomic sequence comparison with Bacillus subtilis.";  
RL Nucleic Acids Res. 28:4317-4331(2000).  
DR EMBL; AP001507; BAB03805.1; -.  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMS; TIGR00671; baf; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 254 AA; 27907 MW; 0E3B557BA7DAC176 CRC64;

Query Match 30.4%; Score 389.5; DB 16; Length 254;  
Best Local Similarity 35.6%; Pred. No. 1.4e-25;  
Matches 90; Conservative 54; Mismatches 88; Indels 21; Gaps 8;  
QY 1 MYLLVDVGNTHSVFSITEDGKTFRRWRRLSTGVFQTEDE-----LFSHLHPLLGDMRE 53  
Db 1 MILVIDVGNNTVLGVYQDETIVHHWRLATSRQKTEDEYAMTVRSLEFDH---AGLQFQD 56  
QY 54 IKGIGVASVVPQTONTVIERFSQKYFHISPIWV--KAKNGCVKWNVK--NPSEVGADRVAN 109  
Db 57 IDGIVISSVPPMMFSLQMKCKKYFHTVPMIIGPGIKTGL---NIKYDNPKEVGADRVN 113  
QY 110 VVAFVKEYGKNGIIIDMGATTVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKLPLVEV 168  
Db 114 AVAAIELYGPPIVVDFTATYCLINEKKQYAGGVIAPGIMISTEALYHRASKLPRIE 173  
QY 169 KPADFVVGKDTENIRLGVNNGSVYALEGIIGRIK-EVYGDLPVWLTGGQSKIIV-KDMIK 226  
Db 174 AKPKQWGTNTIDSMQSGIFYGVSVQVDGVVKRMKAQAESEPKVIATGGLAKLIGTSET 233

227 HEIFDEDLTIKGV 239  
234 IDVIDSFLTLKGL 246

RESULT 3  
Q8YAC5  
ID Q8YAC5 PRELIMINARY; PRT; 259 AA.  
AC Q8YAC5;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical protein lmo0221.  
GN LMO0221.  
OS Listeria monocytogenes.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Listeriaceae; Listeria.  
OX NCBI\_TaxID=1639;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=EGD-E / SEROVAR 1/2A;  
RX MEDLINE=21537279; PubMed=11679669;  
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
RA Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,  
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,  
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,  
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,  
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,  
RA Vazquez-Boland J.-A., Voss H., Wehlend J., Cossart P.;  
RT "Comparative genomics of Listeria species.";  
RL Science 294:849-852(2001).  
DR EMBL; AL591974; CAD00748.1; -.  
DR MEROPS; M41.009; -.  
DR ListiList; LMO00221; -.  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMS; TIGR00671; baf; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 259 AA; 28187 MW; F106B049D80025B7 CRC64;

Query Match 29.5%; Score 378.5; DB 16; Length 259;  
Best Local Similarity 35.6%; Pred. No. 1.3e-24;  
Matches 90; Conservative 52; Mismatches 90; Indels 21; Gaps 7;  
QY 1 MYLLVDVGNTHSVFSITEDGKTFRRWRRLSTGVFQTEDEL-----FSLHPLLGDM 51  
Db 1 MILVIDVGNNTCTGVYKQKLLKHWRMTDRHRTSDELGMTVLNFFSYANLTP----- 54  
QY 52 REIKGIGVASVVPQTONTVIERFSQKYFHISPIWV--KAKNGCVKWNKNPSEVGADRVAN 109  
Db 55 SDIQGIISVVPPIMHAMETMCVRYFNIRPLIVGPGIKTG-LNLKVDNPREIGSDRVN 113  
QY 110 VVAFVKEYGKNGIIIDMGATTVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKLPLVEV 168  
Db 114 AVAASEYGPVIVVDFTATTCYIDESGVYGGAIAPGIMISTEALYNRAAKLPRVDI 173  
QY 169 KPADFVVGKDTENIRLGVNNGSVYALEGIIGRIKEYGDLPPVL-TGOSKIVKDMIKH 227  
Db 174 AESSQIGKSTVSSMQAGIFYGVGQCEGIIAEMKKQSNASPVVATGGLARMITEKSSA 233  
QY 228 -EIFDEDLTIKGV 239  
Db 234 VDILDPFLTLKGL 246

RESULT 4  
Q92F54  
ID Q92F54 PRELIMINARY; PRT; 259 AA.  
AC Q92F54;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical protein lin0253.



Db 68 VEGVISSVVPNIMYSLEHMIRKYFKINPLVVGPIKGTGI---NIKYDNPKEVGADRIYN 124

QY 110 VVAFVKEYGKNGIIDMGTTATVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKLPLVE- 167

Db 125 AVAAHEIYKRSLLIIDFGTATTCFAVRENGDYLGAICPGIKVSSEALFEKAALPRVEL 184

QY 168 VKPADFVVGKDEENIRLGVVNGSVYALEGIIGRIKEYV---GDLP--VVLTTGGQSKIVK 222

Db 185 IKPA-YAICKNTISSIONGIVYIGVYIVVERMKELQEEGEKEPLVVATGGLAKLIS 243

QY 223 DMIKH-EIFDEDLTIKGV 239

Db 244 EEAKNVDVINPFLTLEGL 261

RESULT 7

Q9A621 PRELIMINARY; PRT; 261 AA.

AC Q9A621;

Q 01-JUN-2001 (TReMBLrel. 17, Created)

DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)

DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)

DE Transcriptional activator, putative, Baf family.

GN CC1935.

OS Caulobacter crescentus.

OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;

OC Caulobacter.

OX NCBI\_TaxID=155892;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 19089 / CB15;

RX MEDLINE=21173698; PubMed=11259647;

RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,

RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,

RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,

RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,

RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,

RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,

RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.; }

RT "Complete genome sequence of Caulobacter crescentus.";

RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).

DR EMBL; AE005867; AAK23910.1; -.

DR TIGR; CC1935; -.

DR InterPro; IPR004619; Baf.

DR Pfam; PF03309; Bvg\_acc\_factor; 1.

DR TIGRFAMS; TIGR00671; baf; 1.

KW Complete proteome.

SEQUENCE 261 AA; 27965 MW; C19E60D7B0714EF5 CRC64;

- Query Match 28.1%; Score 359.5; DB 16; Length 261;

Best Local Similarity 35.5%; Pred. No. 5.7e-23;

Matches 89; Conservative 53; Mismatches 94; Indels 15; Gaps 9;

QY 1 MYLLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDELSHLHPLL---GDAMREIKGI 57

Db 2 MLLAIEQGNTMTMFAIHGASWVAQWRSATESRTRTADEYVVVLSQLLSMQGLGFRDAIV 61

QY 58 GVASVVPQTNTVIERFSQKYFHISPIWV--KAKNGCVKWNVPSEVGADRVANVAVFK 115

Db 62 IISVVYPQSIFNLRLNLSRRYFNVEPLVIGENAKLG-IDVRIEKPSEAGADRLVNAIGRAM 120

QY 116 EYKNGIIDMGTTATVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKLPLVEV-KPA-D 172

Db 121 VYGPPLVVIDSGTATTFDIVADGAFEGGIIAPGINLSMQALHEAAKLPRIAIQRPACN 180

QY 173 FVVGKDEENIRLGVVNGSVYALEGIIGRIKEYYGD-LPVVLTGGQSKIVK---DMIKHE 228

Db 181 RIVGTDVTSAMQSGVFWGVIISLIEGLVARIKAERGEPMVTIATGGVASLFEGATDSIDH- 239

QY 229 IFDEDLTIKGV 239

Db 240 -FDSDLTIRGL 249

RESULT 8

Q8XHL5 PRELIMINARY; PRT; 259 AA.

AC Q8XHL5;

DT 01-MAR-2002 (TReMBLrel. 20, Created)

DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)

DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)

DE Hypothetical protein CPE2468.

GN CPE2468.

OS Clostridium perfringens.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;

OC Clostridiales; Clostridiaceae; Clostridium.

OX NCBI\_TaxID=1502;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=13 / TYPE A;

RX PubMed=11792842;

RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,

RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;

RT "Complete genome sequence of Clostridium perfringens, an anaerobic

RT flesh-eater.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).

DR EMBL; AP003194; BAB82174.1; -.

DR InterPro; IPR004619; Baf.

DR InterPro; IPR000515; BPD\_transp.

DR Pfam; PF03309; Bvg\_acc\_factor; 1.

DR TIGRFAMS; TIGR00671; baf; 1.

DR PROSITE; PS00402; BPD\_TRANS\_INN\_MEMBR; UNKNOWN\_1.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 259 AA; 28819 MW; OD5FEA3B7A145E10 CRC64;

Query Match 26.7%; Score 342.5; DB 16; Length 259;

Best Local Similarity 37.6%; Pred. No. 1.6e-21;

Matches 96; Conservative 44; Mismatches 94; Indels 21; Gaps 9;

QY 1 MYLLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDELSHLHPLLGDAM---REIKGI 57

Db 1 MILLIDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDELSHLHPLLGDAM---REIKGI 57

QY 58 GVASVVPQTNTVIERFSQKYFHISPIWV--KAKNGCVKWNVK--NPSEVGADRVANVAV 113

Db 61 IISVVVPNIMHSLNMMVRKCFCKEPIVVGPIKGTGI---NIKYDNPKEVGADRVANVAA 117

QY 114 VKEYGKNGIIDMGTTATVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKLPLVEVKPAD 172

Db 118 FEKHKPMIIDFGTATTCFAITEKGDYILGNICPGIQISADALFERAAKLPRIELEKPK 177

QY 173 FVVGKDEENIRLGVVNGSVYALEGIIGRIKEYYGDL---PVVL-TGGQSKIV---KDM 224

Db 178 SVICKNTVTSMQAGIIGYIGKVEYIVKRMKEMMDLGEKEPFVLATGGLAKLVYSETDV 237

QY 225 IKHEIFDEDLTIKGV 239

Db 238 I--DEVDRKLTLEGL 250

RESULT 9

Q9X8N6 PRELIMINARY; PRT; 265 AA.

AC Q9X8N6;

DT 01-NOV-1999 (TReMBLrel. 12, Created)

DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)

DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)

DE Hypothetical protein SCO3380.

GN SCO3380 OR SCE94.31C.

OS Streptomyces coelicolor.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

OX NCBI\_TaxID=1902;

RN [1]

RP SEQUENCE FROM N.A.



RC STRAIN=A3(2);  
RA Oliver K., Harris D.;  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.;  
RX MEDLINE=97000351; PubMed=8843436;  
RA Kinashi H., Hopwood D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for  
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
RL Mol. Microbiol. 21:77-96(1996).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2) / M145;  
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,  
Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
Hopwood D.A.;  
RT "Complete genome sequence of the model actinomycete Streptomyces  
coelicolor A3(2).";  
RL Nature 417:141-147(2002).  
DR EMBL; AL049628; CAB40880.1; -  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMs; TIGR00671; baf; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 265 AA; 28203 MW; E34FA2ABBA946969 CRC64;

Query Match 26.7%; Score 342.5; DB 16; Length 265;  
Best Local Similarity 35.7%; Pred. No. 1.7e-21;  
Matches 92; Conservative 45; Mismatches 98; Indels 23; Gaps 9;  
QY 1 MYLLVDVGNTHSVESITEDGKTFRRWRLSTGVFQTEDEL-----FSLHPLLLGDAMRE- 53  
DB 1 MLLTIDVGNTHTVLGLFDGEDIVEHWRISTDSRRTADELAVLLQGLMGHPLLGDELGDG 60  
Y 54 IKGIGVASVPTQNTVIERFSQKYFHISPIWVKAKNGCVKNV-----KNPSEVGADRVAN 109  
DB 61 IDGIAICATVPSVLHELREVTTRYGDVPA-VLVEPG-VKTGVPILTDPHPKEVGADRIIN 118  
QY 110 VVAFVKEYKNGIIDMGTTATVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKLPLVEV 168  
DB 119 AVAAVELYGGPAIVDFGTATTFDAVSARGEYIGGVIAPGIEISVEALGVKGAQLRKIEV 178  
QY 169 KPADFVVGKDTENIRLGVNNGSVYALEGIIGRIKEVYGDLP-----VVLGGQSKIVKDM 224  
DB 179 ARPRSVIGKNTVEAMQSGIVYGFAGQVDGVVNRMARLADDPDDVTVIATGGLAPMV--L 236  
QY 225 IKHEIFED-----LTIKGV 239  
DB 237 GESSVIDEHEPWLTLMLGL 254

RESULT 10  
Q8R7M2  
ID Q8R7M2 PRELIMINARY; PRT; 255 AA.  
AC Q8R7M2;  
DT 01-JUN-2002 (TReMBLrel. 21, Created)  
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Putative transcriptional regulator, homologs of Bvg accessory  
factor.

GN TTE2381.  
OS Thermoanaerobacter tengcongensis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.  
OX NCBI\_TaxID=119072;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MB4T / JCM11007;  
RX MEDLINE=21992816; PubMed=11997336;  
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,  
Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,  
Tan H., Chen R., Wang J., Yu J., Yang H.;  
RA "A complete sequence of T. tengcongensis genome.";  
RT Genome Res. 12:689-700(2002).  
RL EMBL; AE013180; AAM25520.1; -  
DR Complete proteome; Hypothetical protein.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 255 AA; 27816 MW; C3C620ECBC8CA6ED CRC64;

Query Match 25.3%; Score 324.5; DB 16; Length 255;  
Best Local Similarity 35.7%; Pred. No. 5.7e-20;  
Matches 89; Conservative 46; Mismatches 101; Indels 13; Gaps 7;  
QY 1 MYLLVDVGNTHSVESITEDGKTFRRWRLSTGVFQTEDE---LPSHLHPLLLGDAMREIKGI 57  
DB 1 MLLAFDVGNNTNIVMGVFKGKLLHSFRISTDKNKTYDEYGLVNLQLYNGISLTEIDDV 60.  
QY 58 GVASVPTQNTVIERFSQKYFHISPIWV--KAKNGCVKNV--NPSEVGADRVANVAF 113  
DB 61 IISVVPPMLNTLQVMSLKYFRTPVIVGPGIKTGI---NIKYDNPKVEGADRIVNAVA 117  
QY 114 VKEYKNGIIDMGTTATVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKLPLVEVKPAD 172  
DB 118 YELYGPPVIVDFGTATTFCAISEKGEYLGIIAPGLMISADALFQRTAKLPKIDLTKPP 177  
QY 173 FVVGKDTENIRLGVNNGSVYALEGIIGRIK-EVYGDLPVLTGGQSKIVKDMIKH-EIF 230  
DB 178 TVINRNTVASMOSGIIYGHVGMVDYIVTRMKGEFAPSAYVATGGFANMAEESKTIDTV 237  
QY 231 DEDLTIKGV 239  
DB 238 NEMLTLEGL 246

RESULT 11  
Q8RFE4  
ID Q8RFE4 PRELIMINARY; PRT; 256 AA.  
AC Q8RFE4;  
DT 01-JUN-2002 (TReMBLrel. 21, Created)  
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Bvg accessory factor.  
GN FN0761.  
OS Fusobacterium nucleatum (subsp. nucleatum).  
OC Bacteria; Fusobacteria; Fusobacterium.  
OX NCBI\_TaxID=76856;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 25586;  
RX MEDLINE=21886394; PubMed=11889109;  
RA Kapral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,  
Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,  
Vasileva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,  
Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,  
Fonstein M., Kyrpides N., Overbeek R.;  
RT "Genome sequence and analysis of the oral bacterium Fusobacterium  
nucleatum strain ATCC 25586.";  
RL J. Bacteriol. 184:2005-2018(2002).  
DR EMBL; AE010586; AAL94957.1; -  
KW Complete proteome.  
SQ SEQUENCE 256 AA; 28601 MW; CDF681127F08914B CRC64;  
Query Match 24.8%; Score 317.5; DB 16; Length 256;  
Best Local Similarity 32.9%; Pred. No. 2.3e-19;

	Matches	82; Conservative	58; Mismatches	100; Indels	9; Gaps	6; Gaps
QY	1	MYLLVDVGNTHSVFSITE-DGKTFRWRRLSTGVFQTEDELSHLPLL---	GDAMREIKG	56		
		: : :       : : :   : : :           : : :				
Db	1	MIIGIDIGNTHIVTGIYDNGELISTFRIATNDKWTEDFYFYFNITKYNEISIKKVDA	60			
		: : :       : : :     : : :   : : :   : : :				
QY	57	IGVASVVPTQNTVIERFSQKYPHISPIWKAKNCGCVKNVK--NPSEVGADRVANVAVFV	114			
		: :       : :   : :     : :   : :   : :   : :				
Db	61	ILISSVVPNIITTFQFARKYFKVEATIVDLEKKLPFTFAKGINYTGFGADRIIDITEAM	120			
		: :       : :   : :     : :   : :   : :   : :				
QY	115	KEY-GKNGIIDIIMGTATTVDLVVNGSYEGGAILPGFFMMVHSLFRGTAKLPLVEVKPADF	173			
		: :     : :           : :           : : :   : :   : :				
Db	121	QKYPDKNLVIFDFGTATTYDVLKKGVIYIGGGILPGIDMSINALYNTAKLPRVKFTTPSS	180			
		: :     : :   : :   : :   : :   : :   : :   : :				
QY	174	VVGKDTEENIRLGVNNGSVVALEGIIGRI-KEYVGDLPVLTGGOSKIVKDMIKH-EIFD	231			
		:     : :   : :   : :   : :   : :   : :   : :				
Db	181	VLGTDTMKQIOAIFFGYAGQIKHIIKKINEELNEEIFVLATGGGLKILSAEIDEIDYD	240			
		: :     : :   : :   : :   : :   : :   : :				
QY	232	EDLTIKGVY	240			
		:   : :     :				
QY	241	ANLSLKGly	249			

RESULT	12
Q9CD56	
ID	Q9CD56 PRELIMINARY; PRT; 274 AA.
AC	Q9CD56;
DT	01-JUN-2001 (TrEMBLrel. 17, Created)
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE	Hypothetical protein ML0232.
GN	ML0232.
OS	Mycobacterium leprae.
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX	NCBI_TaxID=1769;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=TN;
RX	MEDLINE=21128732; Pubmed=11234002;
RA	Cole S.T., Eiglmeler K., Parkhill J., James K.D., Thomson N.R.,
RA	Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA	Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA	Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.
RA	Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA	Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA	Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA	Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA	Barrell B.G.;
RA	"Massive gene decay in the leprosy bacillus.";
RA	Nature 409:1007-1011(2001).
DR	EMBL; AL583917; CAC29740.1; -.
DR	Lepronia; ML0232; -.
DR	InterPro; IPR004619; Baf.
DR	Pfam; PF03309; Bvg_acc_factor; 1.
DR	TIGRFAMS; TIGR00671; baf; 1.
KW	Hypothetical protein; Complete proteome.
SO	SEQUENCE 274 AA, 29421 MW; IC2E735BDEC78765 CRC64;

	Query Match	22.6%;	Score 289;	DB 16;	Length 274;
	Best Local Similarity	32.3%;	Pred. No. 7.2e-17;		
	Matches 83;	Conservative 47;	Mismatches 107;	Indels 20;	Gaps 8;
QY	1	MYLLVDVGNTHSVFSI----	TEDGKTFRRRRLSTGVFQTE	DELFSLHPLGLGDAMREIKG	56
Db	1	MLLAIDVRNTHTVVGLLSGSKEHAKVVQQWRIRTESEVTADELALI	IDGLIGDDSERLAG	60	
QY	57	IGVASVVPTQNTVIERFSQKYFHISP---	IWYKAKNGCVKWNKNPSEVGADRVANVAVF	113	
Db	61	AAALSTVPVSLHEVRIMLDQYWPVSPHVLIEPGVRTG-IPLLDVNPKEVGADRVNCLAA	119		
QY	114	VKEYGKNGIIIDMGTTATTVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKPLPVEVKPAD	172		

D <b>b</b>	120	FHKFGQAAIVVDFGSSICVDVVS	SAKGEFLGGAIAPGVQVSSDAAARSAALRRVELARPR	179
Q <b>Y</b>	173	FWVGKDTTEENIRLGVVNGSVYALEGI	IGR---IKEVYGDL---PVVLTTGGQSKIIVKDM	224
D <b>b</b>	180	SWVGKNTVECMQAGVVFAGLV	DGLVGRMRQDVEEFGDLGNRVAVVATGHTAPLLPE	239
Q <b>Y</b>	225	IKHEI--FDEDLT	IKGV	239
D <b>b</b>	240	L-HTVDHYDRHLTLHGL		255

RESULT 13  
O06282  
ID O06282 PRELIMINARY; PRT; 272 AA.  
AC O06282;  
DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical 29.3 kda protein (Transcriptional activator, putative,  
DE Baf family).  
GN RV3600C OR MTCY07H7B.22 OR MT3706.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence.";  
RL Nature 393:537-544(1998).

Query Match 22.3%; Score 286; DB 16; Length 272;  
Best Local Similarity 31.4%; Pred. No. 1.3e-16;  
Matches 80; Conservative 49; Mismatches 108; Indels 18; Gaps 7;

QY	1	MYLLVDVGNTHSVFSI	---	TEDGKTFRRWRRLSTGV	QFQTEDELFSHLHPL	LLGDAMREIKG	56
Db	1	MLLAIDVRNTHTVVGL	SGMKEHAKVVQQWR	IRTESEVTADELALT	IDGLIGEDSERLTG	60	
QY	57	IGVASVVPTQNTVIER	FSQYFHSIP---	IWVKAKNGCVKWN	KNPSEVGADRVAN	VVAF	113
Db	61	TAALSTVPVSLHEVR	IMLDQWPSVPHVLI	EPGVRTG-IPL	LVDPNKEVGADRI	VNCLAA	119

QY 114 VKEYGKNGIIIDMGATTTVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKLPLVEVKPAD 172  
Db 120 YDRFRKAAIVVDFGSSICVDVVSAGGEFLGAIAPGVQVSSDAAAARSAAALRRVELARPR 179  
QY 173 FVVGKDTENIRLGVVNGSVYALEGIIGRIKE-----VYGDLPVVLTTGGQSKIVKDMIK 226  
Db 180 SVVGKNTVECMQAGAVFGFAGLVVDGLVGRIRREDVSGFSVDHDAIVATGHTAPLLLP- 238  
QY 227 HEI--FDIEDLTIKGV 239  
Db 239 HTVDHYDQHLTLQGL 253

RESULT 14  
O83446 PRELIMINARY: PRT; 273 AA.  
AC O83446;  
DT 01-NOV-1998 (TReMBLrel. 08, Created)  
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
E Hypothetical protein TP0431.  
GN TP0431.  
OS Treponema pallidum.  
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
OX NCBI\_TaxID=160;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NICHOLS;  
RX MEDLINE=98332770; PubMed=9665876;  
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,  
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,  
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,  
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,  
RA McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,  
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,  
RA Venter J.C.;  
RA "Complete genome sequence of Treponema pallidum, the syphilis  
RT spirochete.";  
RL Science 281:375-388(1998).  
DR EMBL; AE001220; AAC65417.1; -.  
DR TIGR; TP0431; -.  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMS; TIGR00671; baf; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 273 AA; 28472 MW; 439C9C77CB598BC0 CRC64;

Query Match 22.2%; Score 285; DB 16; Length 273;  
Best Local Similarity 32.4%; Pred. No. 1.6e-16;  
Matches 84; Conservative 43; Mismatches 104; Indels 28; Gaps 9;

QY 3 LLVDVGNTHSVFSITED--GKTFRR--WRLSTGVFQTEDELFSLHPLLGDMREIKGIG 58  
Db 2 LLIDVGNHVVFGIQGENGRCVRELFRLAPDARKTQDE-----YSLLIHALCERAGVG 56  
QY 59 -----VASVVPTQNTVIERFSQKYFHISPI---WVKAKNGCVKWNKNP--SEVGA 104  
Db 57 RASLRDAFISSVVPVLTCTIADAVAQISGVQVPVFGPWAYEH---LPVRIPEPVRAEIGT 113  
QY 105 DRVANVAVFVKEYGKNGIIIDMGATTTVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKL 163  
Db 114 DLVANAVAAVHFRSACVVVDCGTALTFTAVDGTGLIQGVIAIAPGLRTAVQSLHTGTAQL 173  
QY 164 PLVEVKPADFVVGKDTENIRLGVVNGSVYALEGIIGRI-KEYVGDLPVVLTTGGQSKIVK 222  
Db 174 PLVPLALPDSVLGKDTTHAVQAGVVRGTLFVIRAMIAOCQKELGCRCAAVITGGLSLRFS 233  
QY 223 DMIKHEIFDEDLTIKGVYH 241  
Db 234 SEVDFPPIDAQLTSLGLAH 252

RESULT 15

Q9RX54 PRELIMINARY: PRT; 262 AA.  
AC Q9RX54;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Hypothetical protein DR0461.  
GN DR0461.  
OS Deinococcus radiodurans.  
OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;  
OC Deinococcaceae; Deinococcus.  
OX NCBI\_TaxID=1299;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=R1;  
RX MEDLINE=20036896; PubMed=10567266;  
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,  
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
RA Fraser C.M.;  
RT "Genome sequence of the radioresistant bacterium Deinococcus  
RT radiodurans R1.";  
RL Science 286:1571-1577(1999).  
DR EMBL; AE001905; AAF10040.1; -.  
DR TIGR; DR0461; -.  
DR InterPro; IPR004619; Baf.  
DR InterPro; IPR001220; Lectin\_legB.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMS; TIGR00671; baf; 1.  
DR PROSITE; PS00307; LECTIN\_LEGUME\_BETA; UNKNOWN\_1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 262 AA; 27839 MW; 965EAD2F78785A0 CRC64;

Query Match 20.3%; Score 260; DB 16; Length 262;  
Best Local Similarity 30.6%; Pred. No. 2.1e-14;  
Matches 75; Conservative 48; Mismatches 114; Indels 8; Gaps 7;

QY 3 LLVDVGNTHSVFSITE-DGKTFRRRLSTGVFQTEDELFSLHPLLGDMREIKGIGV-A 60  
Db 7 LAVDIGNTTVLGLADASGALTHTWIRTNREMLPDDLALQLHGLFTLAGAPIPRAAVLS 66  
QY 61 SVVPTQNTVIERFSQKYFHISPIWVAKN-GCVKWNKPNPSEVGADRVANVAVFVKEYG- 118  
Db 67 SVAPPVGENYALAKRHFMDIDAFVSAENLPDVTVELDTPGSGVADRLCNLFGEKYLGG 126  
QY 119 -KNGIIDMGATTTVDLVVNG-SYEGGAILPGFFMMVHSLFRGTAKLPLVEVKPADFVVG 176  
Db 127 LDYAVVVDFGTSTNFDVVVGRRRFLGGILATGAQVSADALFARAACKLPRITLQAPETAIG 186  
QY 177 KDTENIRLGVVNGSVYALEGIIGRIK-EVYGDLPVVLTTGGQSKIVKDMIKH-EIFDEDL 234  
Db 187 KNTVHALQSLVFGYAEWMDGLLRRIARAELPGEAVAVATGGFSRTVQGCICQEIYDETL 246  
QY 235 TIKGV 239  
Db 247 TLRL 251

Search completed: June 24, 2003, 21:59:24  
Job time : 22.0638 secs

**THIS PAGE BLANK (USPTO)**







STREET: 9330 Zionsville Road  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: USA  
ZIP: 46268  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/036,987A  
FILING DATE: 09-MAR-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stuart, Donald R  
REGISTRATION NUMBER: 28,479  
REFERENCE/DOCKET NUMBER: 50,608  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (317)337-4816  
TELEFAX: (317)337-4847  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2152 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-036-987A-3

Query Match 6.8%; Score 89.5; DB 4; Length 2152;  
Best Local Similarity 24.9%; Pred. No. 1.8;  
Matches 73; Conservative 30; Mismatches 113; Indels 77; Gaps 15;  
QY 7 LAVDIGNTTTLGLADASGALHTHTWRIRTNREML-----PDDLALQLHGLFTL----- 54  
Db 1319 VAVEPGEPTDLAQALWGLLRSTQTEHPDRFVLVDVPEPAQLLPALPGVLACGEPQLAL 1378  
QY 55 --AGAPIPRAAVLSS--VAP-PVGENYALALKRHFMDAFVSAE---NLP-----DVTVE 102  
Db 1379 RRGGAHAPRLAGLGSDDVLPVPDGTGWRLEATRPGSLDGLALVDEPTATAPLGDEVR 1438  
QY 103 LDTPGSVGADRL--CNLFGEAEKYLGGLDYAVVVDFGTSTNFDVVGRRRFLG-----GI 154  
Db 1439 MRAAGVNFDRALIALGMYPGVASLGSEGAGVVVETGPGVT--GLAPGDRVGMIPKAFGP 1496  
QY 155 LATG-----AQVSADALFARAAKLPRLITLOAPETAIGKNTVHALQSLGVFGYAE 209  
Db 1497 LAVADHRMVTTRIPAGWSFARAASVPIVFLTA-----YYALVD--L 1534  
QY 210 RRIRAEPLGEAVAV--ATGGFSRTVQGCIEIDYYDETITLRLGLVELWASRSE 260  
Db 1535 AGLR---PGESLLVHSAAGGVGMAAQLARHLG-----AEVYATASE 1573

RESULT 6  
US-09-370-700-3  
Sequence 3, Application US/09370700  
Patent No. 6274350  
GENERAL INFORMATION:  
APPLICANT: Baltz, Richard H  
APPLICANT: Broughton, Mary C  
APPLICANT: Crawford, Kathryn P  
APPLICANT: Madduri, Krishnamurthy  
APPLICANT: Treadway, Patti J  
APPLICANT: Turner, Jan R  
APPLICANT: Waldron, Clive  
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
FILE REFERENCE: 50489 DIV1  
CURRENT APPLICATION NUMBER: US/09/370,700  
CURRENT FILING DATE: 1999-08-09  
EARLIER APPLICATION NUMBER: US 09/36987  
EARLIER FILING DATE: 1998-03-09  
NUMBER OF SEQ ID NOS: 39

SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 3  
LENGTH: 2152  
TYPE: PRT  
ORGANISM: Saccharopolyspora spinosa  
US-09-370-700-3  
Query Match 6.8%; Score 89.5; DB 4; Length 2152;  
Best Local Similarity 24.9%; Pred. No. 1.8;  
Matches 73; Conservative 30; Mismatches 113; Indels 77; Gaps 15;  
QY 7 LAVDIGNTTTLGLADASGALHTHTWRIRTNREML-----PDDLALQLHGLFTL----- 54  
Db 1319 VAVEPGEPTDLAQALWGLLRSTQTEHPDRFVLVDVPEPAQLLPALPGVLACGEPQLAL 1378  
QY 55 --AGAPIPRAAVLSS--VAP-PVGENYALALKRHFMDAFVSAE---NLP-----DVTVE 102  
Db 1379 RRGGAHAPRLAGLGSDDVLPVPDGTGWRLEATRPGSLDGLALVDEPTATAPLGDEVR 1438  
QY 103 LDTPGSVGADRL--CNLFGEAEKYLGGLDYAVVVDFGTSTNFDVVGRRRFLG-----GI 154  
Db 1439 MRAAGVNFDRALIALGMYPGVASLGSEGAGVVVETGPGVT--GLAPGDRVGMIPKAFGP 1496  
QY 155 LATG-----AQVSADALFARAAKLPRLITLOAPETAIGKNTVHALQSLGVFGYAE 209  
Db 1497 LAVADHRMVTTRIPAGWSFARAASVPIVFLTA-----YYALVD--L 1534  
QY 210 RRIRAEPLGEAVAV--ATGGFSRTVQGCIEIDYYDETITLRLGLVELWASRSE 260  
Db 1535 AGLR---PGESLLVHSAAGGVGMAAQLARHLG-----AEVYATASE 1573

RESULT 7  
US-09-428-517-3  
Sequence 3, Application US/09428517  
Patent No. 6251636  
GENERAL INFORMATION:  
APPLICANT: Betlach, Mary C  
APPLICANT: Shah, Sanjay Krishnakant  
APPLICANT: McDaniel, Robert  
APPLICANT: Tang, Li  
TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE  
FILE REFERENCE: 30062-20029.00  
CURRENT APPLICATION NUMBER: US/09/428,517  
CURRENT FILING DATE: 1999-10-28  
EARLIER APPLICATION NUMBER: 60/120,254  
EARLIER FILING DATE: 1999-02-16  
EARLIER APPLICATION NUMBER: 60/106,100  
EARLIER FILING DATE: 1998-10-29  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 3  
LENGTH: 3816  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Recombinant  
OTHER INFORMATION: Oleandolide PKS  
US-09-428-517-3

Query Match 6.7%; Score 89; DB 4; Length 3816;  
Best Local Similarity 25.0%; Pred. No. 5;  
Matches 65; Conservative 33; Mismatches 102; Indels 60; Gaps 13;  
QY 14 TTTVLGLADASGALHTHTWRIRTNREML-----DDLALQLHGLFTLAGAPIPRAAVLSSVA 69  
Db 504 SSTGAGVVDVGSLSVATRSVFHRAVMVGTDLDSMAGSLAGF--AAGGVVP--GVVSGVA 559  
QY 70 PPVGENYALALKRH-----FMIDAFAVSAENLPDVTVELDTPGSGVADRLCNLPFGA 120  
Db 560 PAEGRRVVVFPPGQSQWVGMAAGLLDACPVFAEVAECAAVLDP--LTGWSLVEVLRG 617  
QY 121 EKYLGGLD-----YAVVDFGTSTNF-----DVVGRGRFL-----GGI-LATGAQV 161

Db 618 EAVLGRVDVVPALWAVMVSLARTWRYYGVEPAAVVGHSGEIAAACVAGLSLADGARV 677  
QY 162 SADALFARAALKPRITLQAPETAIGKNTVH-ALQSGLVFGYAEWVDGLLRIRAEALPGEA 220  
Db 678 -----VVLRSRAIARI-----AGGGGMVSVSLPAGRVRTMLDITYGGR-----S 716  
QY 221 VAVATGGFSRTVQICQEI 240  
Db 717 VAAVNGPSSTVVGDAQALD 736

RESULT 8  
US-09-428-517-2  
; Sequence 2, Application US/09428517  
; Patent No. 6251636  
; GENERAL INFORMATION:  
; APPLICANT: Betlach, Mary C.  
; APPLICANT: Shah, Sanjay Krishnakant  
; APPLICANT: McDaniel, Robert  
; APPLICANT: Tang, Li  
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE  
; FILE REFERENCE: 30062-20029.00  
; CURRENT APPLICATION NUMBER: US/09/428.517  
; CURRENT FILING DATE: 1999-10-28  
; EARLIER APPLICATION NUMBER: 60/120,254  
; EARLIER FILING DATE: 1999-02-16  
; EARLIER APPLICATION NUMBER: 60/106,100  
; EARLIER FILING DATE: 1998-10-29  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4150  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant  
US-09-428-517-2

Query Match 6.7%; Score 89; DB 4; Length 4150;  
Best Local Similarity 23.9%; Pred. No. 5.7;  
Matches 62; Conservative 35; Mismatches 104; Indels 58; Gaps 12;  
QY 14 TTTVLGLADASGALHTWRIRTNREMLP-----DDLALQLHGLFTLAGAPIPRAAVLSSVA 69  
Db 3013 SSTAGVVDVVGWLSLVATRSVFEHRAVAVMGTDLDSMAGSLAGF--AAGGWVP--GVVSGVA 3068  
70 PAVGENYALAKRH-----FMIDAFVSAENLPDVTVELDTPGSGVADRLCNLFGA 120  
3069 PAEGRVVFVFPQGSQWVGMAAGLLDACPVFAEVAECAAVLDR--LTGWSLVEVLRG 3126  
QY 121 EKYLGGLD-----YAVVDFGTSTNF-----DVVGRGRRFL-----GGI-LATGAQV 161  
Db 3127 EAVLGRVDVVPALWAVMVSLARTWRYYGVEPAAVVGHSGEIAAACVAGLSLADGARV 3186  
QY 162 SADALFARAALKPRITLQAPETAIGKNTVH-ALQSGLVFGYAEWVDGLLRIRAEALPGEAV 221  
Db 3187 -----VVLRSRAIARIAGGGGMVSVGLS-----AERVMTMLDITYGGRV---SV 3226  
QY 222 AVATGGFSRTVQICQEI 240  
Db 3227 AAVNGPSSTVVGDAQALD 3245

RESULT 9  
US-07-956-700B-111  
; Sequence 111, Application US/07956700B  
; Patent No. 5539092  
; GENERAL INFORMATION:  
; APPLICANT: Robert Haselkorn and Piotr Gornicki  
; TITLE OF INVENTION: Cyanobacterial and Plant Acetyl-CoA  
; TITLE OF INVENTION: Carboxylase

NUMBER OF SEQUENCES: 116  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: 321 No. 5539092th Clark Street  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60610  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy Disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII-DOS  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/956,700B  
FILING DATE: 19921002  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Thomas E. No. 5539092thrup  
REGISTRATION NUMBER: 33,268  
REFERENCE/DOCKET NUMBER: ARCD:058  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 1-312-744-0090  
TELEFAX: 1-312-755-4489  
INFORMATION FOR SEQ ID NO: 111:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 145 amino acids  
TYPE: Amino acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
MOLECULE TYPE: Peptide  
US-07-956-700B-111

Query Match 6.6%; Score 87.5; DB 1; Length 145;  
Best Local Similarity 29.4%; Pred. No. 0.047;  
Matches 37; Conservative 14; Mismatches 60; Indels 15; Gaps 3;  
QY 120 AEKYLGLDYAVVVDFTSTNFDVVGRGRRFLGGILATGAQVSADALFARAALKPRITLQ 179  
Db 19 AEVTLKSDDFELTVRKAVGVNSVVPVVTAPLSGVVSGLP-SAIPIVAHAAPSP----- 72  
QY 180 APETAIGKNTVHALQSGLVFGYAEWVDGLLR-----RAELPGEAVAVATGGFSR 230  
Db 73 SPEPGTSRAADHAVTSSGSGQPGAKIIDQKLAEVASPMVGTFFYRAPAPGEAVFVEVGDRI 132  
QY 231 TVQGIC 236  
Db 133 QQTVC 138

RESULT 10  
US-08-476-537-111  
; Sequence 111, Application US/08476537  
; Patent No. 5756290  
; GENERAL INFORMATION:  
; APPLICANT: Robert Haselkorn and Piotr Gornicki  
; TITLE OF INVENTION: Cyanobacterial and Plant Acetyl-CoA  
; TITLE OF INVENTION: Carboxylase  
; NUMBER OF SEQUENCES: 116  
; CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: 321 No. 5756290th Clark Street  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60610  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy Disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII-DOS  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/476,537



RESULT 11  
 US-08-485-607-111  
 ; Sequence 111, Application US/08485607  
 ; Patent No. 5792627  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Robert Haselkorn and Piotr Gornicki  
 ; TITLE OF INVENTION: Cyanobacterial and Plant Acetyl-CoA  
 ; TITLE OF INVENTION: Carboxylase  
 ; NUMBER OF SEQUENCES: 116  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Arnold, White & Durkee  
 ; STREET: 321 No. 5792627th Clark Street  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: USA  
 ; ZIP: 60610  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy Disk  
 ; COMPUTER: IBM PC Compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: ASCII-DOS  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/485,607  
 ; FILING DATE: 07-JUN-1995  
 ; CLASSIFICATION: 800  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 07/956,700  
 ; FILING DATE: 10/21/92  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Thomas E. No. 5792627thrup  
 ; REGISTRATION NUMBER: 33,268  
 ; REFERENCE/DOCKET NUMBER: ARCD:058  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 1-312-744-0090  
 ; TELEFAX: 1-312-755-4489

Query Match . 6.6%; Score 87.5; DB 2; Length 145;  
Best Local Similarity 29.48; Pred. No. 0.047;



Search completed: June 24, 2003, 21:49:21  
Job time : 9.65383 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:08:46 ; Search time 11.9453 Seconds  
(without alignments)  
2373.336 Million cell updates/sec

Title: US-09-813-453A-8  
Perfect score: 1324  
Sequence: 1 MPAPFLAVDIGNTTTLVGL.....DETLTLRGLVELWASRSEVR 262

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues  
Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA.\*  
1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	ID	Description
1	1324	100.0	262	9	US-09-813-453A-8
2	437.5	33.0	254	9	US-09-813-453A-47
3	428	32.3	258	9	US-09-813-453A-2
4	415.5	31.4	258	9	US-09-813-453A-49
5	414.5	31.3	255	9	US-09-813-453A-7
6	409.5	30.9	265	9	US-09-813-453A-4
7	405.5	30.6	262	9	US-09-813-453A-45
8	404	30.5	272	9	US-09-712-363-276
9	404	30.5	272	9	US-09-813-453A-5
10	396.5	29.9	260	9	US-09-813-453A-51
11	378.5	28.6	233	9	US-09-813-453A-17
12	377	28.5	256	9	US-09-813-453A-55
13	347.5	26.2	258	9	US-09-813-453A-6
14	337	25.5	250	9	US-09-813-453A-3
15	333.5	25.2	219	9	US-09-813-453A-57
16	271	20.5	212	9	US-09-813-453A-59
17	260	19.6	246	9	US-09-813-453A-9
18	253.5	19.1	257	9	US-09-813-453A-53
19	248	18.7	273	9	US-09-813-453A-10

20	215	16.2	241	9	US-09-813-453A-63	Sequence 63, Appl
21	159	12.0	262	9	US-09-813-453A-11	Sequence 11, Appl
22	133.5	10.1	229	9	US-09-813-453A-12	Sequence 12, Appl
23	131	9.9	244	9	US-09-813-453A-41	Sequence 41, Appl
24	128.5	9.7	242	9	US-09-813-453A-65	Sequence 65, Appl
25	123.5	9.3	257	9	US-09-813-453A-13	Sequence 13, Appl
26	104	7.9	505	9	US-09-742-580-8	Sequence 8, Appl
27	104	7.9	505	9	US-09-742-581-8	Sequence 8, Appl
28	104	7.9	505	10	US-09-742-582-8	Sequence 8, Appl
29	103.5	7.8	267	9	US-09-813-453A-15	Sequence 15, Appl
30	101.5	7.7	460	9	US-09-813-453A-39	Sequence 39, Appl
31	100.5	7.6	592	9	US-09-813-453A-22	Sequence 22, Appl
32	99	7.5	984	9	US-10-024-368-2	Sequence 2, Appl
33	99	7.5	1189	9	US-10-024-368-4	Sequence 4, Appl
34	99	7.5	1189	9	US-10-122-013-17	Sequence 17, Appl
35	98	7.4	249	9	US-09-813-453A-70	Sequence 70, Appl
36	96.5	7.3	592	9	US-09-813-453A-43	Sequence 43, Appl
37	95	7.2	249	9	US-09-813-453A-61	Sequence 61, Appl
38	95	7.2	655	9	US-09-906-209-10	Sequence 10, Appl
39	93.5	7.1	317	12	US-10-043-238-1	Sequence 1, Appl
40	93.5	7.1	317	12	US-10-043-238-3	Sequence 3, Appl
41	92	6.9	1189	9	US-10-024-368-3	Sequence 3, Appl
42	91.5	6.9	346	9	US-10-213-878-12	Sequence 12, Appl
43	91.5	6.9	346	9	US-10-214-059-12	Sequence 12, Appl
44	91.5	6.9	872	10	US-09-815-242-13295	Sequence 13295, A
45	90.5	6.8	346	9	US-10-213-878-22	Sequence 22, Appl

ALIGNMENTS

RESULT 1  
US-09-813-453A-8  
; Sequence 8, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 262  
; TYPE: PRT  
; ORGANISM: Deinococcus radiopugnans  
US-09-813-453A-8

Query Match	100.0%	Score 1324;	DB 9;	Length 262;
Best Local Similarity	100.0%	Pred. No. 4.7e-111;		
Matches 262;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MPAPFLAVDIGNTTTLVGLADASGALHTWIRTNREMLPDDDLALQLHGLFTLAGAPIP	60	
Db	1	MPAPFLAVDIGNTTTLVGLADASGALHTWIRTNREMLPDDDLALQLHGLFTLAGAPIP	60	
Qy	61	RAAVLSSVAPPVGENYALALKRHFMDAFVSAENLPDVTVELDTPGSGVADRLCNLFGA	120	
Db	61	RAAVLSSVAPPVGENYALALKRHFMDAFVSAENLPDVTVELDTPGSGVADRLCNLFGA	120	
Qy	121	EKYLGLDYAVVDFGTSTNFDVVGRRFLGGILATGAQVSADALFARAAKLPRLTLOA	180	
Db	121	EKYLGLDYAVVDFGTSTNFDVVGRRFLGGILATGAQVSADALFARAAKLPRLTLOA	180	
Qy	181	PETAIGKNTVHALQSLGVFGYAEVMVDGLLRIRRAELPGEAVATGGFSRTVOGICQEID	240	
Db	181	PETAIGKNTVHALQSLGVFGYAEVMVDGLLRIRRAELPGEAVATGGFSRTVOGICQEID	240	

Db 181 PETAIGKNTVHALQSLVFGYAEMVDGLLRIRRAELPGEAVAVATGGFSRTVQICQEI 240  
QY 241 YYDETTLRLGLVWASRSEVR 262  
Db 241 YYDETTLRLGLVWASRSEVR 262

RESULT 2  
US-09-813-453A-47  
; Sequence 47, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 47  
; LENGTH: 254  
; TYPE: PRT  
; ORGANISM: Bacillus halodurans  
US-09-813-453A-47

Query Match 33.0%; Score 437.5; DB 9; Length 254;  
Best Local Similarity 35.7%; Pred. No. 1.4e-31;  
Matches 89; Conservative 56; Mismatches 97; Indels 7; Gaps 4;  
QY 6 LLAVDIGNTTVLGLADASGALHTHTWRIRTNREMLPDDDLALQLHGLFTLAGAPIP--RAA 63  
Db 2 ILVIDVGNNTVLGVYQ-DETLVHHWRLATSRQKTEDEYAMTVRSLEPDHAGLQFQDIDGI 60  
QY 64 VLSSVAPPVGENYALAKRHFMDAFVSAENLPDVTVELDTPGSGVADRLCNLFGEAKY 123  
Db 61 VISSVPPMMFSLQMKCKYFHVTPMIIGPGIKTGLNLIKYNPNKEVGADRIVNAVAIEL 120  
QY 124 LGGLDY-AVVVDFGTSTNFDVVGRRFLGGILATGAQVSADALFARAAKLPRITLQAPE 182  
Db 121 YG--YPAIVVDFGTATTYCLINEKKQYAGGVIAPGIMISTEALYHRASKLPRIEIAKPK 177  
QY 183 TAIGNTVHALQSLVFGYAEMVDGLLRIRRAELPGEAVAVATGGFSRTVQICQEI 242  
178 QVVGNTIDSMQSGIFGYVSQVDGVVVKMKAQAESEPKVIATGGLAKLIGTSETIDVI 237  
QY 243 DETTLRLGL 251  
Db 238 DSFLLTKGL 246

RESULT 3  
US-09-813-453A-2  
; Sequence 2, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 258  
; TYPE: PRT  
; ORGANISM: Bacillus subtilis  
US-09-813-453A-2  
Query Match 32.3%; Score 428; DB 9; Length 258;  
Best Local Similarity 37.4%; Pred. No. 1e-30;  
Matches 98; Conservative 51; Mismatches 99; Indels 14; Gaps 6;  
QY 6 LLAVDIGNTTVLGLADASGALHTHTWRIRTNREMLPDDDLALQLHGLFTLAGAPIP--AA 63  
Db 2 LLVIDVGNNTVLGVYH-DGKLEWHWRIETSRHKTEDEFGMILRSFLDHSGLMFEQIDGI 60  
QY 64 VLSSVAPPVGENYALAKR----HFMDAFVSAENLPDVTVELDTPGSGVADRLCNLF 119  
Db 61 IISVVVPI----MFALERMCTKYFHIEPQIVGPGMKTGLNLIKYNPNKEVGADRIVNAVA 116  
QY 120 AEKYLGGLDYAVVDFGTSTNFDVVGRRFLGGILATGAQVSADALFARAAKLPRITLQ 179  
Db 117 AIHLYG--NPLIVVDFGTATTYCYIDENKQYMGGAIAPIGTTISTEALYSRAAKLPRIEIT 174  
QY 180 APETAIGNTVHALQSLVFGYAEMVDGLLRIRRAELPGEAVAVATGGFSRTVQICQEI 239  
Db 175 RPDNIIGKNTVSAMQSGILFGYVQVGEIVKRMKQAKQDLKVIATGGLAPLIANESDCI 234  
QY 240 DYYDETTLRLGLVWASRSEV 261  
Db 235 DIVDPFLLTKGL-ELIYERNRV 255

RESULT 4  
US-09-813-453A-49  
; Sequence 49, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 49  
; LENGTH: 258  
; TYPE: PRT  
; ORGANISM: Bacillus stearothermophilus  
US-09-813-453A-49

Query Match 31.4%; Score 415.5; DB 9; Length 258;  
Best Local Similarity 35.2%; Pred. No. 1.3e-29;  
Matches 94; Conservative 49; Mismatches 99; Indels 25; Gaps 6;  
QY 6 LLAVDIGNTTVLGLADASGALHTHTWRIRTNREMLPDDDLALQLHGLFTLAGAPIP--RAA 63  
Db 2 IFVLDVGNNTVLGVYDGD-ELKHHWRIETSRKTEDEYGMKIKALLNHVGLQFSDIRGI 60  
QY 64 VLSSVAPPVGENYALAKR----HFMDAFVSAENLPDVTVELDTPGSGVADRLCN--- 116  
Db 61 IISVVVPI----MFALERMCLKYFHIKPLIVGPGIKTGLDIKYNPREVGADRIVNAVA 116  
QY 117 ---LFGAEKYLGGLDYAVVDFGTSTNFDVVGRRFLGGILATGAQVSADALFARAACL 173  
Db 117 GIHLYGSP-----LIIVDFGTATTYCYINEHKQYMGGAIAPIGIMISTEALFARAACL 168  
QY 174 PRITLQAPETAIGNTVHALQSLVFGYAEMVDGLLRIRRAELPGEAVAVATGGFSRTVQ 233

Db 169 PRIEIAIPDDIIGKNTVSAMQAGILYGVQVEGIVSRMKAQSKIPKVIATGGGLAPLIA 228  
QY 234 GICQEIYDDETLTLRGLVELWASRSE 260  
Db 229 SESDIIDVDPFLTLTGLKLLYEKNT 255

RESULT 5  
US-09-813-453A-7  
; Sequence 7, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 255  
; TYPE: PRT  
; ORGANISM: Geobacter sulfurreducens  
US-09-813-453A-7

Query Match 31.3%; Score 414.5; DB 9; Length 255;  
Best Local Similarity 36.8%; Pred. No. 1.6e-29;  
Matches 96; Conservative 50; Mismatches 102; Indels 13; Gaps 6;  
QY 6 LLAVDIGNTTTLGLADASGALHTWRTNREMLPDDDLALQLHGLFTLAGAPIP--RAA 63  
Db 2 LLVIDVGNNTIVLGIYDGE-RLVRDWRVSTDKARTDEYGILNELFRLAGLGLDQIRAV 60  
QY 64 VLSSVAPP---VGENYALAKRHFMDAFVSAENLPDVTVELDTPGSGVADRLCN-LFG 119  
Db 61 IISVVPLTGVLERLSLG---YFGMRPLVVGPGIKTGMPIQYDNPREVGADRIYNAVAG 117  
QY 120 AEKYLGLDYAVVVDGCTSTNFDVVGRRRFLGGILATGAQVSADALFARAACLPRITLQ 179  
Db 118 YEKYRTSL---IIVDFGTATTFDYVNRKGEYCGGAIAPGLVISTEALFORASKLPRVDII 174  
180 APETAIGKNTVHALQSLVFGYAEWVDGLLRIRAEPLGAEAVAVATGGFSRTVQICQEI 239  
Db 175 RPSAIIARNTVNSMQAGIYYGVGLVDEIVTRMKAESKDAPRVATGGLASLIAPESKTI 234  
QY 240 DYDETLTLRGLVELWASRSE 260  
Db 235 EAVEEYLTLEGLRILYERNRE 255

RESULT 6  
US-09-813-453A-4  
; Sequence 4, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21

; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 265  
; TYPE: PRT  
; ORGANISM: Streptomyces coelicolor  
US-09-813-453A-4  
Query Match 30.9%; Score 409.5; DB 9; Length 265;  
Best Local Similarity 37.9%; Pred. No. 4.8e-29;  
Matches 97; Conservative 44; Mismatches 102; Indels 13; Gaps 4;  
QY 6 LLAVDIGNTTTLGLADASGALHTWRTNREMLPDDDLALQLHGLFTL-----AGAP 58  
Db 2 LLTIDVGNTHTVLGLFDGEDIVEH-WRISTDSRRTADELAVLLQGLMGHMLLGDG 60  
QY 59 IPRAAVLSSVAPPVGENYALAKRHFMDAFVSAENLPDVTVELDTPGSGVADRLCNLF 118  
Db 61 IDGIAICATVPSVLHELREVTTRYGDVPAVLVEPGVKTGVPILTDPKKEVGADRIINAV 120  
QY 119 GAEKYLGLDYAVVVDGCTSTNFDVVGRRRFLGGILATGAQVSADALFARAACLPRITL 178  
Db 121 AAVELYGG--PAIVDFGTATTFDAVSARGEYIGGVIAPGIEISVEALGVKGAKLRKIEV 178  
QY 179 QAPETAIGKNTVHALQSLVFGYAEWVDGLLRIRAEPLGAEAVAVATGGFSRTVQGI 235  
Db 179 ARPRSVIGKNTVEAMQSGIVYFAGQVDGVVNRMARELADDDPDDVTVIATGGLAPMVLGE 238  
QY 236 CQEIDYYDETTLTLRGL 251  
Db 239 SSVIDEHEPWLTLML 254

RESULT 7  
US-09-813-453A-45  
; Sequence 45, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 45  
; LENGTH: 262  
; TYPE: PRT  
; ORGANISM: Bacillus anthracis  
US-09-813-453A-45

Query Match 30.6%; Score 405.5; DB 9; Length 262;  
Best Local Similarity 31.7%; Pred. No. 1.1e-28;  
Matches 84; Conservative 60; Mismatches 104; Indels 17; Gaps 4;  
QY 6 LLAVDIGNTTTLGLADASGALHTWRTNREMLPDDDLALQLHGLFTLAGAPIP--RAA 63  
Db 2 IFVLDVGNTHTVLGVFE-EGELRQHWRMETDRKTEDEYGMVLVKQLLEHGLSFEDVKGI 60  
QY 64 VLSSVAPPVGENYALAKRHFMDAFVSAENLPDVTVELDTPGSGVADRLCN-----L 117  
Db 61 IVSSVPPINFALERMCCEYKIKPLVVGPGIKTGLNIKYENPREVGADRIYNAVAGIHL 120  
QY 118 FGAEKYLGLDYAVVVDGCTSTNFDVVGRRRFLGGILATGAQVSADALFARAACLPRIT 177  
Db 121 YGSP-----LIIVDFGTATTCYCYNEEKHYMGVITPGIMISAEALYSRAAKLPRIE 172







QY	6	LLAVDIGNTTTLVLGLADASGALHTHTWRI	TNR	EMLPDD	LALQLHGLFT	LACAP-IPRAAV	64	
Db	2	LLCIDCGNTNTVFSVMDGTD-FAATWRIATD	HRRTADEYFVWLNTLMQLKGLQGR	ISEAI	60			
QY	65	LSSVAPPVGENYALAKRHFEMIDAF	AVSAE--NL	PDVTVELDTPG	SVGADR	LCN-LFGAE	121	
Db	61	ISSTAPRVFENLRVLCNRYFDCRPV	VVGKPGCELP-VAPRVD	PGTTVGPDR	LVNTVAGYD	119		
QY	122	KYLGGLDYAVVVDFTSTNEDVVGRGR	RFLGGILATGAQVSADALFARA	AKLPRIT	LQAP	181		
Db	120	RHGGDL---IVVDFTATTDFDVVPAD	GAYIGGVIAPGVNLSLEALHMAA	ALPHVDVTKP	176			
QY	182	ETAIGKNTVHALQSLVFGYAEMVDGL	LRRI	RAELPGEAVAVATGFS	NTVQGC	IDEY	241	
Db	177	QGVIGTNTVACIQSGVYWGVI	GLVEGIVRQIRMERDR	PMKV	IATGGLAS	FLDGLG	FDK	236
QY	242	YDETLLT	LRGL	251				
Db	237	VEDDLTMHGL	246					

RESULT 14  
US-09-813-453A-3  
; Sequence 3, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 250  
; TYPE: PRT  
; ORGANISM: Clostridium acetobutylicum  
US-09-813-453A-3

RESULT 15  
US-09-813-453A-57  
; Sequence 57, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers

```

; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGG-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 57
; LENGTH: 219
; TYPE: PRN
; ORGANISM: Dehalococcoides ethenogenes
US-09-813-453A-57

```

Search completed: June 24, 2003, 22:23:56  
Job time : 13.9453 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:33:11 ; Search time 4.55691 Seconds  
(without alignments)  
2384.688 Million cell updates/sec

Title: US-09-813-453A-8  
Perfect score: 1324  
Sequence: 1 MPAFPLLAVDIGNTTTVLGL.....DETTLRLGLVELWASRSEVR 262

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

-Searched: 112892 seqs, 41476328 residues  
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	378.5	28.6	233	1 YACB_BACSU	P37564 bacillus su
2	116	8.8	336	1 Y757_STRCO	Q9rjd5 streptomyc
3	103.5	7.8	267	1 BAF_BORPE	Q45338 bordetella
4	102.5	7.7	675	1 ATKB_DEIRA	Q9rzp0 deinococcus
5	99	7.5	1189	1 HAIR_HUMAN	O43593 homo sapien
6	95	7.2	872	1 SYA_STRPY	Q99257 streptococ
7	94.5	7.1	310	1 KHSE_THEAC	Q9hkr6 thermoplasm
8	94.5	7.1	314	1 MDH_RICCN	Q92ia0 rickettsia
9	93.5	7.1	419	1 AROA_METTH	O26860 methanobact
10	91	6.9	439	1 AROA_HALN1	Q9hgcl halobacteri
11	89.5	6.8	292	1 DAPF_RALSO	Q8y344 ralstonia s
12	89	6.7	383	1 HIS2_NEIMA	Q9jv26 neisseria m
13	88.5	6.7	328	1 GPDA_DEIRA	Q9rr76 deinococcus
14	88.5	6.7	521	1 VL2_HPV04	Q07862 human papil
15	88.5	6.7	1403	1 VG22_HSV11	Q00105 ictalurid h
16	88	6.6	321	1 GLK_BACSU	P54495 bacillus su
17	87.5	6.6	181	1 BCCP_ANASP	Q06881 anabaena sp
18	87.5	6.6	704	1 DP3E_RHOCA	O68045 rhodobacter
19	87.5	6.6	872	1 SYA_STRPN	Q97q48 streptococc
20	87	6.6	365	1 LPXK_CHLPN	Q9z823 chlamydia p
21	87	6.6	383	1 HIS2_NEIMB	Q9k013 neisseria m
22	87	6.6	411	1 VGLM_HSVBC	P52370 bovine herp
23	86.5	6.5	526	1 VP5_BT2A	P30209 bluetongue
24	85.5	6.5	218	1 Y5A4_STRCO	Q9rkms streptomyc
25	85	6.4	603	1 FTSH_CYAME	Q9tj83 cyanidiosch
26	84.5	6.4	928	1 DPO1_SALTY	Q9f173 salmonella
27	84.5	6.4	1025	1 SLAP_CAUCR	P35828 caulobacter
28	84	6.3	360	1 LPXD_CHLPN	Q9z8n6 chlamydia p
29	84	6.3	482	1 PTSB_VIBCH	Q9kvd9 vibrio chol
30	83.5	6.3	348	1 LPXK_AGRF5	Q8uh15 agrobacteri
31	83.5	6.3	375	1 HEM2_STRCO	O50533 streptomyc
32	83.5	6.3	377	1 HIS8_MYCLE	Q9x7b8 mycobacteri
33	83	6.3	499	1 GLCD_ECOLI	P52075 escherichia

34	82.5	6.2	230	1 RNFE_SALTY	Q8xex9 salmonella
35	82.5	6.2	231	1 RNFE_ECOLI	P77179 escherichia
36	82.5	6.2	511	1 DLDH_SCHPO	O00087 schizosacch
37	82.5	6.2	1320	1 PUTA_ECOLI	P09546 escherichia
38	82	6.2	750	1 HEPH_HSV11	P10192 herpes simp
39	81.5	6.2	627	1 FTH1_SYNY3	Q55700 synecocyst
40	81.5	6.2	933	1 SLAP_CAMFE	P35827 campylobact
41	81	6.1	719	1 ATKB_HALN1	P57699 halobacteri
42	81	6.1	835	1 CC48_YEAST	P25694 saccharomyc
43	80.5	6.1	409	1 PEPT_SALTY	P26311 salmonella
44	80.5	6.1	581	1 PBP2_NEIGO	P08149 neisseria g
45	80.5	6.1	644	1 FTSH_ODOSI	P49825 odontella s

ALIGNMENTS

RESULT 1  
YACB\_BACSU  
ID YACB\_BACSU STANDARD; PRT; 233 AA.  
AC P37564;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein yacB.  
GN YACB.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=96051385; PubMed=7584024;  
RA Ogasawara N., Nakai S., Yoshikawa H.;  
RT "Systematic sequencing of the 180 kilobase region of the Bacillus subtilis chromosome containing the replication origin.";  
RL DNA Res. 1:1-14(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G., Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S., Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S., Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M., Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A., Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T., Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D., Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N., Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G., Guisepi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A., Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L., Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C., Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M., Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C., Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H., Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M., Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y., Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F., Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B., Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K., Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A., Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A., Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T., Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K., Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.";  
RL Nature 390:249-256(1997).  
CC -!- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.





RL Mol. Microbiol. 19:37-52(1996).  
CC -|- FUNCTION: ACTIVATES TRANSCRIPTION OF THE PERTUSSIS TOXIN OPERON IN  
CC. A BVGAS-DEPENDENT MANNER. MAY INTERACT WITH THE ALPHA SUBUNIT OF  
CC RNA POLYMERASE.  
CC  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U12020; AAA75361.1; -  
DR EMBL; AF016461; AAC68834.1; -  
DR EMBL; X90711; CAA62242.1; -  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
KW Transcription regulation; Activator.  
? SEQUENCE 267 AA; 27845 MW; 336A615F67B57901 CRC64;

Query Match 7.8%; Score 103.5; DB 1; Length 267;  
Best Local Similarity 24.4%; Pred. No. 0.11;  
Matches 69; Conservative 31; Mismatches 128; Indels 55; Gaps 12;  
QY 6 LLAVDIGNTTTLGLADASGALHTHTWRTNREMLP--DDLALQLHG--LFTLAGAPIP 60  
DB 2 IILIDSGNSRLKVGWFDPPAP-----QAAREPAPVAFDNLDLALGRWLATLPRRP-Q 53  
QY 61 RAAVLSSVAPPVGENYALALK-----RHEMIDAFVSAENLPDVTVELDTPGSVADRL 114  
DB 54 RALGVNVAGLARGEIAATLRAGGCDIRWLRAQPLAMGLRN-----GYRNPDLQAGDRW 107  
QY 115 CNLFGAEKYLGGLDYAVV-DFGTSTNFDVVGRRRFLGGILATGAQVSADALFARAACL 173  
DB 108 ACMVGVLARQPSVHPPLLVASFGTATTLDTIGPDNVFPGGLILPGPAMMRGALAYGTAHL 167  
QY 174 PRITLQAPETAIGKNTVHALQSLVFGVYAEVMDGLLRIRRAELPGE-----AVAVATG 226  
DB 168 PLADGLVADYPI--DTHQAIASGIAAAQAGAI-----VRQWLAGRQRYGQAPEIYVAGG 219  
QY 227 GFSRTVQ-----GICQEIDYDETLTLRGLVELWA 256  
DB 220 GWPEVRQEAERLLAVTGAAGATPQPTLYLDSVP-LDGLAALAA 261

RESULT 4  
TKB\_DEIRA  
ATKB\_DEIRA STANDARD; PRT; 675 AA.  
Q9RZP0;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Potassium-transporting ATPase B chain (EC 3.6.3.12) (Potassium-  
DE translocating ATPase B chain) (ATP phosphohydrolase [potassium-  
DE transporting] B chain) (Potassium binding and translocating subunit  
DE B).  
GN KDPB OR DRB0083.  
OS Deinococcus radiodurans.  
OG Plasmid MPl.  
OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;  
OC Deinococcaceae; Deinococcus.  
OX NCBI\_TaxID=1299;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=R1;  
RX MEDLINE=20036896; PubMed=10567266;  
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,  
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,

RA Fraser C.M.;  
RT "Genome sequence of the radioresistant bacterium Deinococcus  
RT radiodurans R1.";  
RL Science 286:1571-1577(1999).  
CC -|- FUNCTION: ONE OF THE COMPONENTS OF THE HIGH-AFFINITY ATP-DRIVEN  
CC POTASSIUM TRANSPORT (OR KDP) SYSTEM, WHICH CATALYZES THE  
CC HYDROLYSIS OF ATP COUPLED WITH THE EXCHANGE OF HYDROGEN AND  
CC POTASSIUM IONS (BY SIMILARITY).  
CC -|- CATALYTIC ACTIVITY: ATP + H(2)O + K(+)(Out) = ADP + phosphate +  
CC K(+)(In).  
CC -|- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -|- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY  
CC (E1-E2 ATPASES). SUBFAMILY 1A.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AE001826; AAF12535.1; -  
DR TIGR; DRB0083;  
DR InterPro; IPR001757; ATPase\_E1-E2.  
DR InterPro; IPR001454; Hlgase/hydrase.  
DR Pfam; PF00122; E1-E2\_ATPase; 1.  
DR Pfam; PF00702; Hydrolase; 1.  
DR PRINTS; PR00119; CATAPASE.  
DR PROSITE; PS00154; ATPASE\_E1\_E2; 1.  
KW Hydrolase; Potassium transport; Transmembrane; Phosphorylation;  
KW Magnesium; ATP-binding; Plasmid; Complete proteome.  
FT DOMAIN 1 36 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 37 57 POTENTIAL.  
FT DOMAIN 58 63 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 64 82 POTENTIAL.  
FT DOMAIN 83 217 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 218 240 POTENTIAL.  
FT DOMAIN 241 248 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 249 266 POTENTIAL.  
FT DOMAIN 267 566 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 567 586 POTENTIAL.  
FT DOMAIN 587 610 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 611 629 POTENTIAL.  
FT DOMAIN 630 675 CYTOPLASMIC (POTENTIAL).  
FT MOD\_RES 304 304 PHOSPHORYLATION (BY SIMILARITY).  
FT METAL 512 512 MAGNESIUM (BY SIMILARITY).  
FT METAL 516 516 MAGNESIUM (BY SIMILARITY).  
SQ SEQUENCE 675 AA; 70301 MW; 0A033AAB471FC5BB CRC64;  
Query Match 7.7%; Score 102.5; DB 1; Length 675;  
Best Local Similarity 23.3%; Pred. No. 0.4;  
Matches 57; Conservative 38; Mismatches 79; Indels 71; Gaps 12;  
QY 8 AVDIGNTTTLGLADASGALHTHTWRTNREMLPDDLALQLHGLFTLAGAPIPRAAVLSS 67  
DB 291 AVEVAGDVVL-LLDKGTGTT-----IGNRMATKFPPLPGVTEAEALAKAALLSS 338  
QY 68 VAPPVGENYAL-ALKRHFMDAFVSAENLPDVTVELDTPGSVADRLCNLFGAEKYLGG 126  
DB 339 LADPTPEGKSIVALARQLGVDA-----PEPAGAEFI--EFTAQTRMSG 379  
QY 127 LDYAVVDFGTSTNFDVVGRRRFLGGILATGAQVSADALFARAACLPRITLQAPETA- 185  
DB 380 VDFP-----GTS-----IRKG-----AGSRISALAQ-ERGGQLP-----PELAAI 413  
QY 186 -----GKNTVHALQSLVFGVYAEVMDGLLRIRRAELP-----GEAVAVATGGFSRTVQ 233  
DB 414 TDEVSQGGATPLTVIENDRLGLGVVALSDIIPKIRERFERFEQLRRMGLRTVMTGDNPLTAE 473  
QY 234 GICQE 238  
DB 474 AIAKE 478

```
RESULT 5
HAIR_HUMAN
ID HAIR_HUMAN STANDARD; PRT; 1189 AA.
AC O43593; Q9NPE1;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hairless protein.
GN HR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (LONG ISOFORM), AND VARIANT ALUNC ALA-1022.
RX MEDLINE=98111413; PubMed=9445480;
RA Ahmad W., ul Haque M.F., Brancolini V., Tsou H.C., Ul Haque S.,
PA Lam H., Alta V.M., Owen J., Deblaquiere M., Frank J.,
Cserhalmi-Friedman P.B., Leask A., McGrath J.A., Peacocke M.,
Ahmad M., Ott J., Christiano A.M.;
"Alpecia universalis associated with a mutation in the human hairless
gene.";
RL Science 279:720-724(1998).
RN [2]
RP SEQUENCE FROM N.A. (LONG ISOFORM), REVISIONS TO 572 AND 774, AND
RP TISSUE SPECIFICITY.
RC TISSUE=Peripheral blood leukocytes, and Skin fibroblast;
RX MEDLINE=99162400; PubMed=10051399;
RA Ahmad W., Zlotogorski A., Panteleyev A.A., Lam H., Ahmad M.,
RA ul Haque M.F., Abdallah H.M., Dragan L., Christiano A.M.;
RT "Genomic organization of the human hairless gene (HR) and
RT identification of a mutation underlying congenital atrichia in an Arab
RT Palestinian family.";
RL Genomics 56:141-148(1999).
RN [3]
RP SEQUENCE FROM N.A., VARIANT ALUNC ASP-1136, TISSUE SPECIFICITY, AND
RP ALTERNATIVE SPLICING.
RC TISSUE=Peripheral blood leukocytes, Brain, and Fetal brain;
RX MEDLINE=98409496; PubMed=9736769;
RA Cichon S., Anker M., Vogt I.R., Rohleder H., Putzstuck M., Hillmer A.,
RA Farooq S.A., Al-Dhafri K.S., Ahmad M., Haque S., Rietschel M.,
RA Propping P., Kruse R., Noethen M.M.;
RT "Cloning, genomic organization, alternative transcripts and mutational
RT analysis of the gene responsible for autosomal recessive universal
RT congenital alopecia.";
RL Hum. Mol. Genet. 7:1671-1679(1998).
RN [4]
RP VARIANT APL GLN-620.
RX MEDLINE=98431781; PubMed=9758627;
RA Ahmad W., Irvine A.D., Lam H., Buckley C., Bingham E.A.,
RA Panteleyev A.A., Ahmad M., McGrath J.A., Christiano A.M.;
RT "A missense mutation in the zinc-finger domain of the human hairless
RT gene underlies congenital atrichia in a family of Irish travellers.";
RL Am. J. Hum. Genet. 63:984-991(1998).
CC -!- FUNCTION: MAY ACT AS A TRANSCRIPTION FACTOR THAT COULD ACT ON TO
CC REGULATE ONE OF THE PHASES OF HAIR GROWTH.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS: At least 2 isoforms; a long form (shown
CC here) and a short form; are produced by alternative splicing.
CC -!- TISSUE SPECIFICITY: Strongest expression of both isoforms is seen
CC in the small intestine, weaker expression in brain and colon, and
CC trace expression is found in liver, pancreas, spleen, thymus,
CC stomach, salivary gland, appendix and trachea. Long isoform is
CC always the most abundant. Long isoform is exclusively expressed at
CC low levels in kidney and testis and short isoform exclusively at
CC high levels in the skin.
CC -!- DISEASE: DEFECTS IN HR ARE THE CAUSE OF ALOPECIA UNIVERSALIS
CC (ALUNC); A RARE AUTOSOMAL RECESSIVE FORM OF HAIR LOSS. IT IS
CC CHARACTERIZED BY HAIR FOLLICLES WITHOUT HAIR.
CC -!- DISEASE: DEFECTS IN HR ARE THE CAUSE OF ATRICHIA WITH PAPULAR
CC LESIONS (APL) (ALSO KNOWN AS CONGENITAL ATRICHIA). THIS AUTOSOMAL
```

```
CC RECESSIVE DISEASE IS CHARACTERIZED BY PAPILLARY LESIONS OVER MOST
CC OF THE BODY AND ALMOST COMPLETE ABSENCE OF HAIR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF039196; AAC32258.2; -
DR EMBL; AJ277249; CAB87577.2; -
DR EMBL; AJ277250; CAB87577.2; JOINED.
DR EMBL; AJ277251; CAB87577.2; JOINED.
DR EMBL; AJ277252; CAB87577.2; JOINED.
DR EMBL; AJ277253; CAB87577.2; JOINED.
DR EMBL; AJ400825; CAB87577.2; JOINED.
DR EMBL; AJ400826; CAB87577.2; JOINED.
DR EMBL; AJ400827; CAB87577.2; JOINED.
DR EMBL; AJ400828; CAB87577.2; JOINED.
DR EMBL; AJ400829; CAB87577.2; JOINED.
DR EMBL; AJ400830; CAB87577.2; JOINED.
DR EMBL; AJ400831; CAB87577.2; JOINED.
DR EMBL; AJ400832; CAB87577.2; JOINED.
DR EMBL; AJ400833; CAB87577.2; JOINED.
DR EMBL; AJ400834; CAB87577.2; JOINED.
DR EMBL; AJ400835; CAB87577.2; JOINED.
DR EMBL; AJ400836; CAB87577.2; JOINED.
DR EMBL; AJ400837; CAB87577.2; JOINED.
DR EMBL; AJ277165; CAB86602.1; -
DR Genew; HGNC:5172; HR.
DR MIM; 602302; -
DR MIM; 203655; -
DR MIM; 209500; -
DR InterPro; IPR003347; TF_JmJc.
DR Pfam; PF02373; JmJc; 1.
KW Zinc-finger; DNA-binding; Nuclear protein; Transcription regulation;
KW Metal-binding; Alternative splicing; Disease mutation.
FT ZN_FING 600 625
FT VARSPLIC 1072 1126
FT VARIANT 620 620
FT VARIANT 1022 1022
FT VARIANT 1136 1136
FT CONFLICT 337 337
FT CONFLICT 446 446
FT CONFLICT 584 584
FT SEQUENCE 1189 AA; 127509 MW; 5E244858716EB5DF CRC64;
Query Match 7.5%; Score 99; DB 1; Length 1189;
Best Local Similarity 25.6%; Pred. No. 1.5;
Matches 69; Conservative 24; Mismatches 107; Indels 70; Gaps 13;
QY 2 PAFPLLAVDIGNTTTVLGLA-----DASGALTH-----TWIRTNREMLPDDLA 45
DB 903 PLGPPQPSSLSGTFWEGFSWPELRPKSDEGSVLLHLRALGDEDTSRVENLAASLPLEY 962
QY 46 LQLHGLFTLAGAPIPRAAVLSSVAPPVGVENYALALKR-HFMIDAFVSAENLPDVTVELD 104
DB 963 CALHGKLNLA-SYLPPLGALRPLEPQLWAAAYGVSPHRGHLGTKNLCVEADLVSLVHAD 1021
QY 105 TPGSVGADRLCNLFGAEK-YLGGLDYAVVDFGTFSTNFDVVGRRRFLGGILATGAQVSA 163
DB 1022 TP-----LPAWHRAQKDFLSGLD-----GE-----GLWSPGSGYST 1052
QY 164 DALFARAAKLPRI-----TLQAPETAIGKNTVHAIQSLGVLFGYAEVMDGLLRRAE---- 215
DB 1053 VWHVFRAQDAQRIRRFLOMVCAPAG-----GALEPGAP-GSCYLDAGLRRRLREWGVS 1105
QY 216 -----LPGEAVALVATGGFSRTVQIGCQEI 239
```

Db	1106	CWTLQAPGEAVLPAGA-PHQVQGLVSTV	1134			1134
RESULT 6						
SYA_STRPY	1106	CWTLQAPGEAVLPAGA-PHQVQGLVSTV	1134			1134
ID	SYA_STRPY	STANDARD;	PRT;	872	AA.	
AC	Q99257;					
DT	15-JUN-2002	(Rel. 41, Created)				
DT	15-JUN-2002	(Rel. 41, Last sequence update)				
DT	15-JUN-2002	(Rel. 41, Last annotation update)				
DE	Alanyl-tRNA synthetase (EC 6.1.1.7)	(Alanine--tRNA ligase) (Alars).				
GN	ALAS OR SPY1389.					
OS	Streptococcus pyogenes.					
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;					
OC	Streptococcus.					
OX	NCBI_TaxID=1314;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
STRAIN=SF370 / ATCC 700294 / Serotype M1;						
MEDLINE=21192684; PubMed=11296296;						
RA	Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,					
RA	Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,					
RA	Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,					
RA	Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;					
RT	"Complete genome sequence of an M1 strain of Streptococcus pyogenes.";					
RL	Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).					
CC	-1- CATALYTIC ACTIVITY: ATP + L-alanine + tRNA(Ala) -> AMP +					
CC	diphosphate + L-alanyl-tRNA(Ala).					
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic.					
CC	-1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.					
CC	-----					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration					
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -					
CC	the European Bioinformatics Institute. There are no restrictions on its					
CC	use by non-profit institutions as long as its content is in no way					
CC	modified and this statement is not removed. Usage by and for commercial					
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/					
CC	or send an email to license@isb-sib.ch).					
CC	-----					
DR	EMBL; AE006576; AAK34208.1; ..					
DR	InterPro; IPR002106; AATRNA_ligaseII.					
DR	InterPro; IPR003156; DHHA1.					
DR	InterPro; IPR002318; tRNA-synt_2c.					
DR	Pfam; PF02272; DHHA1; 1.					
DR	Pfam; PF01411; tRNA-synt_2c; 1.					
DR	PRINTS; PR00980; TRNASYNTHALA.					
DR	TIGRFAMS; TIGR00344; alas; 1.					
DR	PROSITE; PS50860; AA_TRNA_LIGASE_II_ALA; 1.					
DR	Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;					
KW	Complete proteome.					
SQ	SEQUENCE 872 AA; 96530 MW; E4605AB8B195B959 CRC64;					
Query Match 7.2%; Score 95; DB 1; Length 872;						
Best Local Similarity 23.5%; Pred. No. 2.3;						
Matches 77; Conservative 35; Mismatches 98; Indels 118; Gaps 17;						
QY	21	ADASGALHTWTRIRNREMLPDDALALQLHGLFTLAGAPIPRAAVLSSVAPPVGENVALAL	80			
Db	501	AEMGGQVADTGRIKNDK---GDTVAEVVDVQKAPNGQPLHTVNVLASLS--VGTNTLEI	555			
QY	81	KR-----HFMIDAF-----AVSAENLPDV	99			
Db	556	NKERRLAVEKNHTATHLHAALHNVI GEHATQAGSLNEEFRLFDFTHFEAVSNEELRHI	615			
QY	100	TVE-----LDTPGSVGADRLCNLFGAKEY-----LGGLDYAVVVD	134			
Db	616	EQEVNEQIWNALTIITTTEDVETAKEMGA---MALFG-EKYGKVVVVQIG--NYSVELC	669			
QY	135	FGTSTN-----FDVV-----GRGRRFLGGILATGAQV-----SADALFARAACLPR	176			
Db	670	GGTHLNNSSEIGLFKIVKEGIGSGTRRI--IAVTGROAFEAYRNQEDALKETIAA-----	722			
Query Match 7.1%; Score 94.5; DB 1; Length 310;						
Best Local Similarity 22.6%; Pred. No. 0.76;						
Matches 62; Conservative 44; Mismatches 87; Indels 81; Gaps 16;						
QY	42	DDLALQLHGLFTLAGAPIP-----RAAVLSSVAPPVGENYALALKRHPMIDAFVSAENL	96			
Db	33	DRVTISDHGDFKLTGEGIPLDPERNTAGLAAIA-----LLDELGL--ERD	75			
QY	97	PDVTVELDTP-----GSVGADRLCNLFGAKEYLG-GL-----DYAVVVDFTSTNFDV	144			
Db	76	IEIRIEKGVPKGLGLGSSGASAAAAYKALDAYFGLGLTQEEMIGYAMIGEASSGSPHPD	135			
QY	145	GRGRRFLGGIL-----ATGAQVSADALF-----ARAAK--LPR-ITLQA	180			
Db	136	NVSASITGGVLVTVTHDDDLRASRINISGDFRFLVAIPDLFTENKTRARQMLPRSIPLDS	195			



QY 181 PETAIGKNTVHALQSLVFGYAEV-----DGLLR--RI-----RAELPGEAVAV 223  
Db 196 YSKALGRTA--SLIAGLMSGNRNLIRVGNDDIVEPSRISLFPYYDKRMALANEAVAA 253  
QY 224 ATGGFSRTVQIGICQIDYDETLTLR-GLVELWA 256  
Db 254 AVSGAGPSILMVCDEMSDMD---SIRAGISEIFS 284

RESULT 8

MDH\_RICCN  
ID MDH\_RICCN STANDARD; PRT; 314 AA.  
AC Q92IA0;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Malate dehydrogenase (EC 1.1.1.37).  
GN MDH OR RC0520.  
AS Rickettsia conorii.  
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
Rickettsiaceae; Rickettsieae; Rickettsia.  
NCBI\_TaxID=781;  
[1]  
SEQUENCE FROM N.A.  
STRAIN=Malish 7;  
MEDLINE=21442074; PubMed=11557893;  
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,  
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,  
RA Raoult D.;  
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";  
RL Science 293:2093-2098(2001).  
CC -!- CATALYTIC ACTIVITY: (S)-malate + NAD(+) -> oxaloacetate + NADH.  
CC -!- SIMILARITY: BELONGS TO THE LDH FAMILY. MDH SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AE008614; AAL03058.1; -  
DR InterPro; IPR001252; Mdh.  
DR InterPro; IPR001236; ldh.  
DR Pfam; PF00056; ldh; 1.  
DR Pfam; PF02866; ldh\_C; 1.  
PROSITE; PS00068; MDH; FALSE\_NEG.  
Oxidoreductase; Tricarboxylic acid cycle; NAD; Complete proteome.  
ACT\_SITE 150 PROTON-RELAY (BY SIMILARITY).  
FT BINDING 153 SUBSTRATE CARBOXYL GROUP (BY SIMILARITY).  
FT ACT\_SITE 177 PROTON-RELAY (BY SIMILARITY).  
SQ SEQUENCE 314 AA; 33638 MW; 27F016CDE3EF6F14 CRC64;

Query Match 7.1%; Score 94.5; DB 1; Length 314;  
Best Local Similarity 23.3%; Pred. No. 0.78;  
Matches 74; Conservative 42; Mismatches 96; Indels 105; Gaps 15;

QY 16 TVLGLADASGALHTHTWRIR-----TNREMLPDDLALQLHGLFTLAG----- 56  
Db 8 SLICSGNIGGTLAHLISLRELGDIVLFDVTEGVPOGKALDLMQAGTIAGSDIKIKGTNDY 67  
QY 57 -----APIPRAAVLS-----SVAPPVGENYALAKRHFMDIDAFVSAENLPDV 99  
Db 68 KDIEGSDAIIITAGLPRKPGMSREDLISINTGIMKTVAANVKK-YAPDAFVIVITNPLDV 126  
QY 100 TV-----ELDTP-----GSYGADRLCNLFGAEKY-----LGGLDYAVV--- 132  
Db 127 MYVYMLKESGLPHNKKVIGMAGVLDSSRF-NLFLAEFEKVSNNVNSVMVLGGHGDAMVPLA 185  
QY 133 -----VDFGTSTNFD---VVGRGRRFLGGILA---TGAQVSADALFARAA-- 171

Db 186 RYSTISGVPIPDLIKMLSSNENIEKIIDTRNGGGEIVALLKTGS-----AYYAPAASA 240  
QY 172 -----KLPRITLQAPETAIGKNTVHALQSLVFGYAEVMDGLLRIRRAELPGEAVAV 223  
Db 241 IEMLESYLKDKRQILTCAAHLQGEYGVHDLVGVPIMIGK--EGVLRVIELQLTAEEKAL 298  
QY 224 ATGGFSRTVQIGICQEID 240  
Db 299 ----FDKSVGEVKKLIE 311

RESULT 9

ARO\_A\_METH  
ID ARO\_A\_METH STANDARD; PRT; 419 AA.  
AC O26860;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Probable 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19)  
DE (5-enolpyruvylshikimate-3-phosphate synthase) (EPSP synthase) (EPSPS).  
GN AROA OR MTH766.  
OS Methanobacterium thermoautotrophicum.  
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;  
OC Methanobacteriiaecae; Methanothermobacter.  
OX NCBI\_TaxID=187420;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Delta H;  
RX MEDLINE=98037514; PubMed=9371463;  
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,  
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,  
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,  
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,  
RA Jiواني N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,  
RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,  
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;  
RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
RT deltaH: functional analysis and comparative genomics.";  
RL J. Bacteriol. 179:7135-7155(1997).  
CC -!- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate -  
CC phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.  
CC -!- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;  
CC sixth step.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
CC -!- SIMILARITY: BELONGS TO THE EPSP SYNTHASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AE000855; AAB85269.1; -  
DR InterPro; IPR001986; EPSP\_synthase.  
DR Pfam; PF00275; EPSP\_synthase; 1.  
DR ProDom; PD001867; EPSP\_synthase; 1.  
DR PROSITE; PS00104; EPSP\_SYNTHASE\_1; 1.  
DR PROSITE; PS00885; EPSP\_SYNTHASE\_2; 1.  
KW Aromatic amino acid biosynthesis; Transference; Complete proteome.  
SQ SEQUENCE 419 AA; 44747 MW; 395BDB8C5262AC9 CRC64;

Query Match 7.1%; Score 93.5; DB 1; Length 419;  
Best Local Similarity 25.5%; Pred. No. 1.3;  
Matches 70; Conservative 38; Mismatches 101; Indels 65; Gaps 14;

QY 11 IGNT-----TTVLGLADASGALHTHTWRIRTNREMLP-----DDLALQLHGLFT 53  
Db 87 LGNSGTTLRMTSVAGLAENYTVLTGDESURT-RPMQPLLDALRPLGVEALSSRMNGLPP 145  
QY 54 L-----AGAPIPRAAVLSS-----VAPPVGENYALAKRHFMDIDAFVSAENLPDVT 100



Db 146 IIVRGLRGSTIRGDVSSQFISSILIAAPLITEGVEVMVEGDFISRPV-----DMT 198

QY 101 VELDTPGSVGADRLCNLFGAEKYL-GGLDYAVVDFGTSTNEDVVGRRFLGGILATGA 159

Db 199 VDVMERFSVPVDYSEGTFRVEPAVYRGLDYTVEGDYSSAS-----YLAGAVAAAGG 249

QY 160 QVSADALFARAANKLPRIITL-----QAPETAIGKNTVHALQSLVFGYAEMVDGLLRIRA 214

Db 250 DVLIENLFRDSRQGDRIILDIISDMGAEVRRGEDHVRIASTGELSG-----VSVNLHDAPD 305

QY 215 ELPGEAV--AVATGGFSRT-VOGICQEIYDET 245

Db 306 LLPTVAVLGALATG---RTEIGGV--EHARYKET 334

RESULT 10

AROQ\_HALN1 STANDARD; PRT; 439 AA.

AC Q9HQC1;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Probable 3-phosphohikimate 1-carboxyvinyltransferase (EC 2.5.1.19)

DE (5-enolpyruvylshikimate-3-phosphate synthase) (EPSP synthase) (EPSPS).

GN AROA OR PSC OR VNGI232G.

OS Halobacterium sp. (strain NRC-1).

OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;

OC Halobacteriaceae; Halobacterium.

OX NCBI\_TaxID=64091;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20504483; PubMed=11016950;

RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,

RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,

RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,

RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,

RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,

RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,

RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,

RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;

RT "Genome sequence of Halobacterium species NRC-1.";

RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).

CC -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate -

CC phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.

CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;

CC sixth step.

CC -1- SUBUNIT: Monomer (By similarity).

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).

CC -1- SIMILARITY: BELONGS TO THE EPSP SYNTHASE FAMILY.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

CC EMBL; AE005049; AAG19594.1; -

DR InterPro; IPR001986; EPSP\_synthase.

DR Pfam; PF00275; EPSP\_synthase; 1.

DR ProDom; PD001867; EPSP\_synthase; 2.

DR PROSITE; PS00104; EPSP\_SYNTHASE\_1; 1.

DR PROSITE; PS00885; EPSP\_SYNTHASE\_2; 1.

KW Aromatic amino acid biosynthesis; Transferase; Complete proteome.

FT ACT\_SITE 439 AA; 44333 MW; 48CAD75A1F0ECB89 CRC64;

FT SEQUENCE

Query Match 6.9%; Score 91; DB 1; Length 439;

Best Local Similarity 23.4%; Pred. No. 2.3;

Matches 64; Conservative 31; Mismatches 85; Indels 94; Gaps 13;

QY 2 PAPPLLAVDIGNTTTVLGLADASGALHTWRIRTNREMLPDDALQLHG--LFTLAG--- 56

||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

Db 87 PAIPDAVIDCANSGTMTMLVTAALADGTTVLTGDE-----SLRAPHGPLLDSGLGG 142

QY 57 -----APIPRAAVLS--SVAPP--VGENYALAKRHFEMIDAFVSAENLP----- 97

Db 143 TARSTRNGCAPLVVDGPVSGGSVALPGDVSSQFVTAL-----LMAGAVTETGIETDLTT 197

QY 98 -----DVTVELDTPGSVGADRLC---NLFGAEKYL-GGLDYAVVDFGTSTNEDVVG 145

Db 198 ELKSAPYVDITLDVDAFGVGASETAAGYRVRGQAYAPSGAEYAVPGDFSSASY----- 252

QY 146 RGRRLGGLATGAQVVSADALFARAANKLPRIITLQAPETAIGKNTVHALQSLVFGYAEMV 205

Db 253 -----LLAAGALAAADG-----AAVVEGMHPSAQ-----GDAAIIV 283

QY 206 D-----GLLRIRRAELPGEAVAVA 224

Db 284 DVLERMGADIDWDTESGVITVQRSELSGVEVGVA 317

RESULT 11

DAPE\_RALSO STANDARD; PRT; 292 AA.

ID DAPE\_RALSO

AC Q8Y344;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Diaminopimelate epimerase (EC 5.1.1.7) (DAP epimerase).

GN DAPF OR RSC0137 OR RS01049.

OS Ralstonia solanacearum (Pseudomonas solanacearum).

OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;

OC Ralstonia.

OX NCBI\_TaxID=305;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=GM11000;

RX MEDLINE=21681879; PubMed=11823852;

RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,

RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,

RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,

RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,

RA Sigulier P., Thebault P., Whalen M., Wincker P., Levy M.,

RA Weissenbach J., Boucher C.A.;

RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";

RL Nature 415:497-502(2002).

CC -1- CATALYTIC ACTIVITY: LL-2,6-diaminoheptanedioate = meso-

CC diaminohexanedioate.

CC -1- PATHWAY: SIXTH STEP IN THE BIOSYNTHESIS OF LYSINE FROM ASPARTATE

CC SEMIALDEHYDE.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

CC -1- SIMILARITY: BELONGS TO THE DIAMINOPIMELATE EPIMERASE FAMILY.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

CC EMBL; AL646057; CAD13665.1; -

DR InterPro; IPR001653; DAP\_epimerase.

DR Pfam; PF01678; DAP\_epimerase; 2.

DR TIGRFAMS; TIGR00652; DapF; 1.

DR PROSITE; PS01326; DAP\_EPIMERASE; FALSE\_NEG.

KW Isomerase; Lysine biosynthesis; Complete proteome.

FT ACT\_SITE 75 75 BY SIMILARITY.

FT ACT\_SITE 226 226 BY SIMILARITY.

FT SEQUENCE 292 AA; 31628 MW; BB6CC03A2939CC51 CRC64;

Query Match 6.8%; Score 89.5; DB 1; Length 292;

Best Local Similarity 23.0%; Pred. No. 1.9;

Matches 65; Conservative 23; Mismatches 103; Indels 91; Gaps 13;

```

47 QY QLHGLFTLAGAPIPRAAVLSSVAPPVG---ENYALAKKRHEMIDA--FAVSAENLPDVT 100
Db :|| || || :|| :|| :|| :|| :|| :|| :||
7 KMHG---AGNDF---VVLDGIATPIDFTPEQWRAIADRHFGVGADQLLVERSTRPDVD 59

101 QY VELDTPGSVGADRLCNLFGAEKYLGGLDYAVVWDEGTSTNFDVVGRGR-----RFLG 152
Db :|| :|| || || || || || || || || ||
60 FRY-----RIFNHDGGEVEQCGNGARCFVKF-----VDRGLTDKRTIRVEVMN 103

153 QY GI---LATGAQVSAD--ALFARAAKLPRITLQAPETAIGKNTVHALQ----- 194
Db :|| :|| || || || || || || || || ||
104 GISTLTMQPDGQVTVDMGAPVFEEAARLPFVPDALTPTRVEGRDTHALQINGRTAWLSTVS 163

195 QY -----SGLVF-----GYAEMVD--GLLRIRRAELPGEA 220
Db :|| :|| :|| :|| :|| :|| :|| :||
164 MGNPHAVQVVDAAEFVREDGCLIESHAVFPRRVNAGFEMETADRHAIRLRVYERGAGET 223

221 QY VAVATGGFSRTVQIGCEIDYDETTLTRG--LVLEWASRSE 260
Db :|| :|| :|| :|| :|| :|| :|| :||
224 LACGTGACAAAAGIRRGLLDSPVKVTHGGDLTIAWAGEGE 265

```

**RESULT 12**

HISZ\_NEIMA STANDARD; PRT; 383 AA.  
 Q9JV26;  
 15-JUN-2002 (Rel. 41, Created)  
 15-JUN-2002 (Rel. 41, Last sequence update)  
 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE ATP phosphoribosyltransferase regulatory subunit.  
 GN HISZ OR NMA1023.  
 OS *Neisseria meningitidis* (serogroup A).  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=65699;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Z2491 / Serogroup A / Serotype 4A;  
 RX MEDLINE=20222556; PubMed=10761919;  
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,  
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,  
 RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,  
 RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,  
 RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,  
 RA Whitehead S., Spratt B.G., Barrell B.G.;  
 RT "Complete DNA sequence of a serogroup A strain of *Neisseria*  
 RT *meningitidis* Z2491.";  
 RL Nature 404:502-506(2000).  
 CC -!- FUNCTION: May allow the regulation of ATP  
 CC phosphoribosyltransferase activity by histidine (By similarity).  
 CC -!- PATHWAY: Histidine biosynthesis; first step. Very important in the  
 CC regulation of histidine metabolism.  
 CC -!- SUBUNIT: Binds to hisG (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
 CC -!- MISCELLANEOUS: This function is generally fulfilled by the C-  
 CC terminal part of hisG, which is missing in some bacteria such as  
 CC this one.  
 CC -!- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.  
 CC HISZ SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collabora-  
 CC tion between the Swiss Institute of Bioinformatics and the EMBL outstat-  
 CC ion at the European Bioinformatics Institute. There are no restrictions on  
 CC use by non-profit institutions as long as its content is in no  
 CC modified and this statement is not removed. Usage by and for commer-  
 CC cial entities requires a license agreement (See <http://www.isb-sib.ch/anno>)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AL162754; CAB84292.1; -.  
 DR Histidine biosynthesis; Complete proteome.  
 KW SEQUENCE 383 AA; 41837 MW; B7BCA7266045F860 CRC64;  
 SQ

QY	9	VDIGNTTTVLGLADASGALHTHTWIRTNREMLPDDLALQLHGLFTLAGAPIP	60
Db.	176	LDAGQSATLLALMQDKDTETVEAQVK-----ANKLGMWAKAFSLPRLYGGREV	225
QY	61	-----RAAVLSSVAPPGENYALALKRHF MIDAFVSAENLPDVTVELDTPGSVGADRL	114
Db	226	LSDARGRLPDL SAVGGALGELQAVC-----DAF-----PDCEIHID-----	261
QY	115	CNLFGEAKYLGLDYAV-VVDFGTSTNFDVVGRRRF--LGGIL-----ATGAQVSADAL	166
Db	262	LSELRVNDNYHTGLLYAAYAADF-----HDAVARGGRYDGLGGYFGRARPATGFSFDLRSE	316
QY	167	FARAAKL---PRITLQAPETAIGKNTVHALQSLVFGYAEMVD-GLLRRIRRAELPG	218
Db	317	IGRLPAIEROPAVLVDAEDAEAAHAHEAVEALREQ---GOCVVYDYGIGHNVSEELAG	369

RESULT 13

```

GPDA_DEIRA
ID GPDA_DEIRA STANDARD; PRT; 328 AA.
AC Q9RR76;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Glycerol-3-phosphate dehydrogenase [NAD(P)+] (EC 1.1.1.94) (NAD(P)H-
DE dependent glycerol-3-phosphate dehydrogenase).
GN GPDA OR DR2621.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1;
RX MEDLINE=20036896; Pubmed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
CC -!- CATALYTIC ACTIVITY: Sn-glycerol 3-phosphate + NAD(P)(+) =
CC glycerone phosphate + NAD(P)H.
CC -!- PATHWAY: De novo phospholipid biosynthesis; glycerol-3 phosphate
CC formation.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- SIMILARITY: BELONGS TO THE NAD-DEPENDENT GLYCEROL-3-PHOSPHATE
CC DEHYDROGENASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE002091; AAF12158.1; -.
CC TIGR; DR2621; -.
DR InterPro; IPR001652; NAD_Gly3p_dh.
DR Pfam; PF01210; NAD_Gly3P_dh; 1.
DR PRINTS; PR00077; GPDHGRNASE.
DR ProDom; PD001649; NAD_Gly3P_dh; 1.
DR PROSITE; PS00957; NAD_G3PDH; 1.
KW Phospholipid biosynthesis; Oxidoreductase; NAD; Complete proteome.
SQ SEQUENCE 328 AA; 33597 MW; 46D909D7D14170BE CRC64;
Query Match 6.7%; Score 88.5; DB 1; Length 328;

```







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:35:06 ; Search time 21.3688 Seconds  
(without alignments)  
2526.317 Million cell updates/sec

Title: US-09-813-453A-8  
Perfect score: 1324  
Sequence: 1 MPAPPLAVIDIGNTTTVLGL.....DETLTLRGLVELWASRSEVR 262

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues 671580  
otal number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL\_21.\*
- 1: sp\_archaea.\*
  - 2: sp\_bacteria.\*
  - 3: sp\_fungi.\*
  - 4: sp\_human.\*
  - 5: sp\_invertebrate.\*
  - 6: sp\_mammal.\*
  - 7: sp\_mhc.\*
  - 8: sp\_organelle.\*
  - 9: sp\_phage.\*
  - 10: sp\_plant.\*
  - 11: sp\_rodent.\*
  - 12: sp\_virus.\*
  - 13: sp Vertebrate.\*
  - 14: sp\_unclassified.\*
  - 15: sp\_rvirus.\*
  - 16: sp\_bacteriap.\*
  - 17: sp\_archaeap.\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1324	100.0	262	16 Q9RX54	Q9rx54 deinococcus
2	437.5	33.0	254	16 Q9KGH5	Q9kgh5 bacillus ha
3	414.5	31.3	255	16 Q8R7M2	Q8r7m2 thermoanaer
4	412.5	31.2	258	2 Q9F985	Q9f985 bacillus st
5	409.5	30.9	265	16 Q9X8N6	Q9x8n6 streptomyce
6	404	30.5	272	16 O06282	O06282 mycobacteri
7	401	30.3	274	16 Q9CD56	Q9cd56 mycobacteri
8	396.5	29.9	261	16 Q9A6Z1	Q9a6z1 caulobacter
9	385.5	29.1	259	16 Q8XHL5	Q8xhl5 clostridium
10	371.5	28.1	273	16 Q97EB4	Q97eb4 clostridium
11	363.5	27.5	259	16 Q8YAC5	Q8yac5 listeria mo
12	356.5	26.9	259	16 Q92F54	Q92f54 listeria in
13	344.5	26.0	256	16 Q8RFE4	Q8rfe4 fusobacteri
14	271	20.5	212	2 O32514	O32514 desulfovibr
15	260	19.6	246	16 Q9WZY5	Q9wzy5 thermotoga
16	248	18.7	273	16 O83446	O83446 treponema p

17	159	12.0	262	16	O51477	O51477 borrelia bu
18	152.5	11.5	276	16	O8YQD7	O8yqd7 anabaena sp
19	139.5	10.5	295	16	O8Y2M4	O8y2m4 ralstonia s
20	133.5	10.1	229	16	O67753	O67753 aquifex aeo
21	128.5	9.7	242	16	O9PC14	O9pci4 xylella fas
22	123.5	9.3	257	16	P74045	P74045 synechocyst
23	107	8.1	438	17	O8TS02	O8ts02 methanosarc
24	104.5	7.9	382	16	O8XUP0	O8xup0 ralstonia s
25	104.5	7.9	400	16	O86549	O86549 streptomyce
26	104	7.9	505	3	Q9Y799	Q9y799 candida par
27	103.5	7.8	255	16	Q9L1E6	Q9l1e6 streptomyce
28	101.5	7.7	444	16	Q9A5G5	Q9a5g5 caulobacter
29	101	7.6	435	16	O92LG6	O92lg6 rhizobium m
30	101	7.6	636	2	Q53818	Q53818 streptomyce
31	101	7.6	636	16	O88039	O88039 streptomyce
32	101	7.6	1441	16	O9CFL1	O9cfl1 lactococcus
33	100.5	7.6	592	16	Q9JWI7	Q9jwi7 neisseria m
34	100	7.6	615	2	Q9X6I8	Q9x6i8 streptomyce
35	99.5	7.5	544	17	Q9YEG2	Q9yeg2 aeropyrum p
36	99.5	7.5	1363	2	Q9F5K8	Q9f5k8 rhodobacter
37	98.5	7.4	302	16	P96939	P96939 mycobacteri
38	98.5	7.4	477	16	Q9L2I1	Q9l2i1 streptomyce
39	97	7.3	323	16	Q99YU9	Q99yu9 streptococc
40	96.5	7.3	515	16	Q8YDQ5	Q8ydq5 brucella me
41	96.5	7.3	592	16	Q9JXF1	Q9jxf1 neisseria m
42	95	7.2	297	2	Q93FF7	Q93ff7 acinetobact
43	95	7.2	2065	2	Q93DC7	Q93dc7 yersinia en
44	95	7.2	2691	16	Q8XP01	Q8xpul ralstonia s
45	95	7.2	11096	2	Q9L4W3	Q9l4w3 streptomyce

ALIGNMENTS

RESULT 1

Q9RX54	ID	Q9RX54	PRELIMINARY;	PRT;	262 AA.
AC	Q9RX54;				
DT	01-MAY-2000	(TREMBLrel. 13, Created)			
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)			
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)			
DE	Hypothetical protein DR0461.				
GN	DR0461.				
OS	Deinococcus radiodurans.				
OC	Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;				
OC	Deinococcaceae; Deinococcus.				
OX	NCBI_TaxID=1299;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=R1;				
RX	MEDLINE=20036896; PubMed=10567266;				
RA	White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,				
RA	Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,				
RA	Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,				
RA	Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,				
RA	Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,				
RA	Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,				
RA	Fraser C.M.;				
RT	"Genome sequence of the radioresistant bacterium Deinococcus				
RT	radiodurans R1."				
RL	Science 286:1571-1577(1999).				
DR	EMBL; AE001905; AAF10040.1; -.				
DR	TIGR; DR0461; -.				
DR	InterPro; IPR004619; Baf.				
DR	InterPro; IPR001220; Lectin_legB.				
DR	Pfam; PF03309; Bvg_acc_factor; 1.				
DR	TIGRFAMS; TIGR00671; baf; 1.				
DR	PROSITE; PS00307; LECTIN_LEGUME_BETA; UNKNOWN_1.				
KW	Hypothetical protein; Complete proteome.				
SQ	SEQUENCE 262 AA; 27839 MW; 965EAD2F78785A0 CRC64;				

Query Match 100.0%; Score 1324; DB 16; Length 262;  
Best Local Similarity 100.0%; Pred. No. 2.5e-95;

Matches 262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 MPAPLLAVDIGNTTTVLGLADASGALHTWRTNRNREMLPDDLALQLHGLFTLAGAPIP 60
Db	1 MPAPLLAVDIGNTTTVLGLADASGALHTWRTNRNREMLPDDLALQLHGLFTLAGAPIP 60
QY	61 RAAVLSSVAPPVGENYALAKRHFMDAFVSAENLPDVTVELDTPGSVGADRLCNLFGA 120
Db	61 RAAVLSSVAPPVGENYALAKRHFMDAFVSAENLPDVTVELDTPGSVGADRLCNLFGA 120
QY	121 EKYLGGLDYAVVDFGTSTNFDVVGRRRFLGGILATGAQVSADALFARAAKLPRTILQA 180
Db	121 EKYLGGLDYAVVDFGTSTNFDVVGRRRFLGGILATGAQVSADALFARAAKLPRTILQA 180
QY	181 PETAIGKNTVHALQSLVFGYAEMVDGLLRIRAEPLGCEAVAVATGGFSRTVQGICQEID 240
Db	181 PETAIGKNTVHALQSLVFGYAEMVDGLLRIRAEPLGCEAVAVATGGFSRTVQGICQEID 240
QY	241 YYDETILTRLGLVELWASRSEVR 262
Db	241 YYDETILTRLGLVELWASRSEVR 262
RESULT 2	
Q9KGH5	
ID	Q9KGH5 PRELIMINARY; PRT; 254 AA.
AC	Q9KGH5
DT	01-OCT-2000 (Tremblrel. 15, Created)
DT	01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE	Hypothetical protein BH0086.
OS	Bacillus halodurans.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC	Bacillaceae; Bacillus.
OX	NCBI_TaxID=86665;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=C-125 / JCM 9153;
RX	MEDLINE=20512582; PubMed=11058132;
RA	Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA	Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA	Horikoshi K.;
RT	*Complete genome sequence of the alkaliphilic bacterium Bacillus
RT	halodurans and genomic sequence comparison with Bacillus subtilis.";
RL	Nucleic Acids Res. 28:4317-4331(2000).
DR	EMBL; AP001507; BAB03805.1; -
DR	InterPro; IPR004619; Baf.
DR	Pfam; PF03309; Bvg_acc_factor; 1.
DR	TIGRFAMs; TIGR00671; baf; 1.
DR	Hypothetical protein; Complete proteome.
W	SEQUENCE 254 AA; 27907 MW; 0E3B557BA7DAC176 CRC64;
SQ	
Query Match 33.0%; Score 437.5; DB 16; Length 254;	
Best Local Similarity 35.7%; Pred. No. 2.4e-26;	
Matches 89; Conservative 56; Mismatches 97; Indels 7; Gaps 4;	
QY	6 LLAVDIGNTTTVLGLADASGALHTWRTNRNREMLPDDLALQLHGLFTLAGAPIP--RAA 63
Db	2 ILVIDVGNNTVLGVQ-DETLVHHWRLATSRQKTEDEYAMTVRSFLDHAGLQFQDIDGI 60
QY	64 VLSSVAPPVGENYALAKRHFMDAFVSAENLPDVTVELDTPGSVGADRLCNLFGAEKY 123
Db	61 VISSVPPPMFSLQMKCKKHYFHTPMIIGPGIKTGLNPKYDNPKEVGADRVNAVAIEL 120
QY	124 LGGLDY-ADVDFGTSTNFDVVGRRRFLGGILATGAQVSADALFARAAKLPRTILQAPE 182
Db	121 YG---YPAIVDFGTATYCLINEKKQYAGGVIAPIAGMISTEALYHRASKLPRIEIAKPK 177
QY	183 TAIGKNTVHALQSLVFGYAEMVDGLLRIRAEPLGCEAVAVATGGFSRTVQGICQEIDYY 242
Db	178 QWVGNTIDMSQSGIFYGVVSQVDGVVKRMAQAESEPKVIATGGLAKLIGTSETIDVI 237

QY	243 DETLRLGL 251
Db	238 DSFLTLKGL 246
RESULT 3	
Q8R7M2	
ID	Q8R7M2 PRELIMINARY; PRT; 255 AA.
AC	Q8R7M2;
DT	01-JUN-2002 (Tremblrel. 21, Created)
DT	01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE	Putative transcriptional regulator, homologs of Bvg accessory
DE	factor.
GN	TTE2381.
OS	Thermoanaerobacter tengcongensis.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC	Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX	NCBI_TaxID=119072;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=MB4T / JCM11007;
RX	MEDLINE=21992816; PubMed=11997336;
RA	Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA	Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA	Tan H., Chen R., Wang J., Yu J., Yang H.;
RT	*A complete sequence of T. tengcongensis genome.";
RL	Genome Res. 12:689-700(2002).
DR	EMBL; AE013180; AAM25520.1; -
KW	Complete proteome; Hypothetical protein.
SQ	SEQUENCE 255 AA; 27816 MW; C3C620ECBC8CA6ED CRC64;
Query Match 31.3%; Score 414.5; DB 16; Length 255;	
Best Local Similarity 35.8%; Pred. No. 1.5e-24;	
Matches 92; Conservative 48; Mismatches 112; Indels 5; Gaps 3;	
QY	6 LLAVDIGNTTTVLGLADASGALHTWRTNRNREMLPDDLALQLHGLFTLAGAPIPRA--A 63
Db	2 LLAFDVGNNTNIVMGVFKGK-KLLHSFRISTDKNKTYDEYGMVLNQLIGYNGISLTEIDDV 60
QY	64 VLSSVAPPVGENYALAKRHFMDAFVSAENLPDVTVELDTPGSVGADRLCNLFGAEKY 123
Db	61 IISVVPPPLMNTLQVMSLKIFRTKPIVVGPGIKTGINIKYDNPKEVGADRVNAVAAYEL 120
QY	124 LGGLDYAVVDFGTSTNFDVVGRRRFLGGILATGAQVSADALFARAAKLPRTILQAPE 183
Db	121 YGG--PVIVIDFGTATTFCAISERGEYLGIIAPGLMISADALFQRTAKLPKIDLTKPPT 178
QY	184 AIGKNTVHALQSLVFGYAEMVDGLLRIRAEPLGCEAVAVATGGFSRTVQGICQEIDYY 243
Db	179 VINRNTVASMOSGIYGHVGMVDYIVTRMKGEFAPSAYVAVATGGFANMAEESKTIDTVN 238
QY	244 ETLRLGLVELWASRSE 260
Db	239 EMLTLEGLRIIYERNKE 255
RESULT 4	
Q9F985	
ID	Q9F985 PRELIMINARY; PRT; 258 AA.
AC	Q9F985;
DT	01-MAR-2001 (Tremblrel. 16, Created)
DT	01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE	Putative 32 kDa replication protein.
OS	Bacillus stearothermophilus.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC	Geobacillus.
OX	NCBI_TaxID=1422;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=V;
RA	Vasquez C., Pichuanes S., Saavedra C.;









RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
RA Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,  
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,  
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
RA Entian K.-D., Fsihl H., Garcia-del Portillo F., Garrido P.,  
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,  
RA Madoeno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,  
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,  
RT "Comparative genomics of *Listeria species.*";  
RL Science 294:849-852(2001).  
DR EMBL; AL591974; CAD00748.1; -.  
DR MEROPS; M41.009; -.  
DR ListiList; LMO00221; -.  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMS; TIGR00671; baf; 1.  
DR Hypothetical protein; Complete proteome.  
SQ SEQUENCE 259 AA; 28187 MW; F106B049D80025B7 CRC64;  
  
Query Match 27.5%; Score 363.5; DB 16; Length 259;  
Best Local Similarity 31.7%; Pred. No. 1.4e-20;  
Matches 80; Conservative 53; Mismatches 114; Indels 5; Gaps 4;  
  
QY 6 LLAVDIGNTTTLGLADASGALHTHTWRIRTNREMLPDDLALQLHGLFTLAG-APIP-RAA 63  
Db 2 ILVIDVGNCTVGVYKEQKLLKH-WRMTDRHRTSDELGMTVNLFFSYANLTPSDIQGI 60  
  
QY 64 VLSSVAPPVGENYALALKRHFMDAFVSAENLPDVTVELDTPGSGVADRLCNLFGAEKY 123  
Db 61 IISVVPPIMHAMETMCVRYFNIRPLIVPGIKTGLNLKVDNPREIGSDRIVNAVAASEE 120  
  
QY 124 LGGLDYAVVDFGTSTNFDVVGRRRFLGGILATGAQVSADALFARAAKLPRITLOAPET 183  
Db 121 YG--TPVIVVDFGTATTTCYIDESGVYQGGAIAPGIMISTEALYNRAAKLPRVDIAESSQ 178  
  
QY 184 AIGKNTVHALQSLGVFGYAEVMDGLLRIRAEPLGAEVAVATGGFSRTVQGICQEIDYYD 243  
Db 179 IIGKSTVSSMQAGIFYGFVQCGEGIIAEMKKQSNASPVVAVATGGLARMITEKSSAVDILD 238  
  
QY 244 ETLTLRGLVELW 255  
Db 239 PFLTGLGLELLY 250  
  
RESULT 12  
Q92F54 PRELIMINARY; PRT; 259 AA.  
Q92F54;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical protein lin0253.  
GN LIN0253.  
OS *Listeria innocua*.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Listeriaceae; Listeria.  
OX NCBI\_TaxID=1642;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CLIP 11262 / SEROVAR 6A;  
RX PubMed=11679669;  
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
RA Baquero F., Bloecker H., Brandt P., Chakraborty T.,  
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,  
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
RA Entian K.-D., Fsihl H., Garcia-del Portillo F., Garrido P.,  
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,  
RA Madoeno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,

RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,  
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,  
RT "Comparative genomics of *Listeria species.*";  
RL Science 294:849-852(2001).  
DR EMBL; AL596164; CAC95486.1; -.  
DR ListiList; LIN00253; -.  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMS; TIGR00671; baf; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 259 AA; 28227 MW; 554B03A0C0EFA64F CRC64;  
  
Query Match 26.9%; Score 356.5; DB 16; Length 259;  
Best Local Similarity 31.7%; Pred. No. 5e-20;  
Matches 80; Conservative 52; Mismatches 115; Indels 5; Gaps 4;  
  
QY 6 LLAVDIGNTTTLGLADASGALHTHTWRIRTNREMLPDDLALQLHGLFTLAG-APIP-RAA 63  
Db 2 ILVIDVGNCTVGVYKEQKLLRH-WRMTDRHRTSDELGMTVNLFFSYANLTPSDIQGI 60  
  
QY 64 VLSSVAPPVGENYALALKRHFMDAFVSAENLPDVTVELDTPGSGVADRLCNLFGAEKY 123  
Db 61 IISVVPPIMHAMETMCVRYFNIRPLIVPGIKTGLNLKVDNPREIGSDRIVNAVAASEE 120  
  
QY 124 LGGLDYAVVDFGTSTNFDVVGRRRFLGGILATGAQVSADALFARAAKLPRITLOAPET 183  
Db 121 YG--TPVIVVDFGTATTTCYIDAGVYQGGAIAPGIMISTEALYNRAAKLPRVDIAESSQ 178  
  
QY 184 AIGKNTVHALQSLGVFGYAEVMDGLLRIRAEPLGAEVAVATGGFSRTVQGICQEIDYYD 243  
Db 179 IIGKSTVASMQAGIFYGFVQCGEGIIAEMKKQSNTPVVAVATGGLARMITEKSSAVDILD 238  
  
QY 244 ETLTLRGLVELW 255  
Db 239 PFLTGLGLELLY 250  
  
RESULT 13  
Q8RFE4 PRELIMINARY; PRT; 256 AA.  
Q8RFE4;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Bvg accessory factor.  
GN FN0761.  
OS *Fusobacterium nucleatum* (subsp. *nucleatum*).  
OC Bacteria; Fusobacteria; Fusobacterium.  
OX NCBI\_TaxID=76856;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 25586;  
RX MEDLINE=21886394; PubMed=11889109;  
RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,  
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,  
RA Vasieva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,  
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,  
RA Fonstein M., Kyrpides N., Overbeek R.;  
RT "Genome sequence and analysis of the oral bacterium *Fusobacterium*  
nucleatum strain ATCC 25586.";  
RL J. Bacteriol. 184:2005-2018(2002).  
DR EMBL; AE010586; AAL94957.1; -.  
KW Complete proteome.  
SQ SEQUENCE 256 AA; 28601 MW; CDF681127F08914B CRC64;  
  
Query Match 26.0%; Score 344.5; DB 16; Length 256;  
Best Local Similarity 32.3%; Pred. No. 4.3e-19;  
Matches 82; Conservative 55; Mismatches 110; Indels 7; Gaps 5;  
  
QY 6 LLAVDIGNTTTLGLADASGALHTHTWRIRTNREMLPDDLALQLHGLFTLAGAPIPR--AA 63  
Db 2 IIGIDIGNTHIVTGIYDNNGELISTFRIATNDKMTDEYFNFNITKYNEISIKKVDAL 61



**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:26 ; Search time 26.899 Seconds  
(without alignments)  
1297.879 Million cell updates/sec

Title: US-09-813-453A-8  
Perfect score: 1324  
Sequence: 1 MPAFPLLAVDIGNTTVLGL.....DETTLRLGLVELWASRSEVR 262

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

---searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1324	100.0	262	AAU91155	Deinococcus radiop
2	437.5	33.0	254	AAU91171	Pantothenate kinas
3	428	32.3	258	AAU01243	B. subtilis novel
4	428	32.3	258	AAU91149	Bacillus subtilis
5	415.5	31.4	258	AAU91172	Pantothenate kinas
6	414.5	31.3	255	AAU91154	Geobacter sulfurre
7	409.5	30.9	265	AAU91151	Streptomyces coeli
8	405.5	30.6	262	AAU91170	Pantothenate kinas
9	404	30.5	272	AAG81225	Mycobacterium tube
10	404	30.5	272	AAU91152	Mycobacterium tube

11	396.5	29.9	260	23	AAU911173	Pantothenate kinas
12	378.5	28.6	233	23	AAU911163	Pantothenate kinas
13	377	28.5	256	23	AAU911175	Pantothenate kinas
14	363.5	27.5	259	23	ABB47661	Listeria monocytog
15	347.5	26.2	258	23	AAU91153	Rhodobacter capsul
16	337	25.5	250	23	AAU91150	Clostridium acetob
17	333.5	25.2	219	23	AAU91176	Pantothenate kinas
18	271	20.5	212	23	AAU91177	Pantothenate kinas
19	260	19.6	246	23	AAU91156	Thermotoga maritim
20	253.5	19.1	257	23	AAU91174	Pantothenate kinas
21	248	18.7	273	23	AAU91157	Treponema pallidum
22	215	16.2	241	23	AAU91179	Pantothenate kinas
23	159	12.0	262	23	AAU91158	Borrelia burgdorfe
24	133.5	10.1	229	23	AAU91159	Aquifex aeolicus p
25	131	9.9	244	23	AAU91168	Pantothenate kinas
26	128.5	9.7	242	23	AAU91180	Pantothenate kinas
27	123.5	9.3	257	23	AAU91160	Synechocystis pant
28	107	8.1	457	22	AAU56858	Propionibacterium
29	104	7.9	505	19	AAW70516	Candida parapsilos
30	104	7.9	505	22	AAG64143	Candida parapsilos
31	103.5	7.8	267	23	AAU91162	Bordella pertussis
32	101.5	7.7	455	20	AAU38617	Neisseria gonorrhoe
33	101.5	7.7	455	21	AAU74908	Neisseria gonorrhoe
34	101.5	7.7	460	23	AAU91167	Pantothenate kinas
35	101.5	7.7	592	20	AAU38618	Neisseria gonorrhoe
36	101.5	7.7	592	21	AAU74911	Neisseria gonorrhoe
37	101	7.6	1440	23	ABB54801	Lactococcus lactis
38	100.5	7.6	455	21	AAU74910	Neisseria meningit
39	100.5	7.6	592	20	AAU38616	Neisseria meningit
40	100.5	7.6	592	21	AAU74913	Neisseria meningit
41	100.5	7.6	592	23	AAU91166	Pantothenate kinas
42	99	7.5	984	23	AAE19794	Human Hairless pro
43	99	7.5	1189	20	AAU15217	Human Hairless wil
44	99	7.5	1189	23	AAE19796	Human Hairless pro
45	98	7.4	249	23	AAU911182	Pantothenate kinas

ALIGNMENTS

RESULT 1	
AAU91155	
ID	AAU91155 standard; Protein; 262 AA.
XX	
AC	AAU91155;
XX	
DT	05-JUN-2002 (first entry)
XX	
DE	Deinococcus radiopugnans pantothenate kinase Coax.
XX	
KW	Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW	pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX	
OS	Deinococcus radiopugnans.
XX	
PN	WO200216601-A2.
XX	
PD	28-FEB-2002.
XX	
PF	24-AUG-2001; 2001WO-US26531.
XX	
PR	24-AUG-2000; 2000US-227860P.
PR	20-MAR-2001; 2001US-0813453.
XX	
PA	(OMNI-) OMNIGENE BIOPRODUCTS INC.
XX	
PI	Yocum RR, Patterson TA;
XX	
DR	WPI; 2002-269358/31.
XX	
PT	Identifying potential antibiotic or antimicrobial agent, comprises
PT	contacting composition comprising pantothenate kinase (Coax) protein
PT	with test compound and identifying inhibitor of the Coax protein -

XX Claim 10; Page 73-74; 128pp; English.

CC The invention describes assays for identifying a (potential) antibiotic

CC comprising contacting an assay composition comprising a pantothenate

CC kinase (Coax) protein with a test compound, and determining the ability

CC of the test compound to inhibit the activity of the Coax protein, an

CC essential enzyme for the production of coenzyme A. Coax protein is a

CC valuable target for identifying bactericidal compounds. Coax modulating

CC agents can be used in an infectious animal model to determine the

CC efficacy, toxicity, or side effects of treatment with such an agent. This

CC is the amino acid sequence of a pantothenate kinase (Coax) protein

CC described in the invention.

SQ Sequence 262 AA;

Query Match 100.0%; Score 1324; DB 23; Length 262;

Best Local Similarity 100.0%; Pred. No. 6.3e-131;

Matches 262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MPAPFLLAVDIGNTTTVLGLADASGALHTWRTNREMLPDDDLALQLHGLFTLAGAPIP 60

1 MPAPFLLAVDIGNTTTVLGLADASGALHTWRTNREMLPDDDLALQLHGLFTLAGAPIP 60

61 RAAVSSVAPPVGENYALALKRHFMDAFVSAENLPDVTVELDTPGSGADRLCNLFGA 120

61 RAAVSSVAPPVGENYALALKRHFMDAFVSAENLPDVTVELDTPGSGADRLCNLFGA 120

121 EKYLGGLDYAVVVDFGTSTNFDVVGRRFLGGILATGAQVSADALFARAAKLPRLITLQA 180

121 EKYLGGLDYAVVVDFGTSTNFDVVGRRFLGGILATGAQVSADALFARAAKLPRLITLQA 180

181 PETAIGNKTVHALQSLVFGYAEMVDGLRRRAELPGEAVAVATGGFSRTVQICQIEID 240

181 PETAIGNKTVHALQSLVFGYAEMVDGLRRRAELPGEAVAVATGGFSRTVQICQIEID 240

241 YYDETLLRGLVELWASRSEVR 262

241 YYDETLLRGLVELWASRSEVR 262

RESULT 2

AAU91171

ID AAU91171 standard; Protein; 254 AA.

XX AC AAU91171;

XX 05-JUN-2002 (first entry)

DT Pantothenate kinase (Coax) #9.

Pantothenate kinase; Coax; antibiotic; antimicrobial;

pantothenate kinase modulator; coenzyme A; bactericidal compound.

Bacillus halodurans.

WO200216601-A2.

28-FEB-2002.

24-AUG-2001; 2001WO-US26531.

24-AUG-2000; 2000US-227860P.

20-MAR-2001; 2001US-0813453.

(OMNI-) OMNIGENE BIOPRODUCTS INC.

Yocum RR, Patterson TA;

WPI; 2002-269358/31.

N-PSDB; ABK54192.

Identifying potential antibiotic or antimicrobial agent, comprises

PT contacting composition comprising pantothenate kinase (Coax) protein

PT with test compound and identifying inhibitor of the Coax protein

XX Claim 10; Page 100; 128pp; English.

XX The invention describes assays for identifying a (potential) antibiotic

CC comprising contacting an assay composition comprising a pantothenate

CC kinase (Coax) protein with a test compound, and determining the ability

CC of the test compound to inhibit the activity of the Coax protein, an

CC essential enzyme for the production of coenzyme A. Coax protein is a

CC valuable target for identifying bactericidal compounds. Coax modulating

CC agents can be used in an infectious animal model to determine the

CC efficacy, toxicity, or side effects of treatment with such an agent. This

CC is the amino acid sequence of a pantothenate kinase (Coax) protein

CC described in the invention.

XX SQ Sequence 254 AA;

Query Match 33.0%; Score 437.5; DB 23; Length 254;

Best Local Similarity 35.7%; Pred. No. 1.4e-37;

Matches 89; Conservative 56; Mismatches 97; Indels 7; Gaps 4;

6 LLAVDIGNTTTVLGLADASGALHTWRTNREMLPDDDLALQLHGLFTLAGAPIP--RAA 63

2 ILVIDVGNTNTVLGVYQ--DETLVHHWRLATSRQKTEDEYAMTVRSFLDHAGLQFQDIDGI 60

64 VLSSVAPPVGENYALALKRHFMDAFVSAENLPDVTVELDTPGSGADRLCNLFGAEKY 123

61 VISSVPPMFMFSLQCKKYYFHVTPMIIGPGIKTGLNPKYDNPKEVGADRVNAVAIEL 120

124 LGGLDY-AVVVDFGTSTNFDVVGRRFLGGILATGAQVSADALFARAAKLPRLITLQAPE 182

121 YG---YPAIVVDFGTATYCLINEKKQYAGGVIAPGIMISTEALYHRASKLPRIEIAKPK 177

183 TAIGNKTVHALQSLVFGYAEMVDGLRRRAELPGEAVAVATGGFSRTVQICQIEIDYY 242

178 QVVGNTIDSMQSGIFYGVSVQDGVVVKRMAQAESEPKVIATGGLAKLIGTESITDVI 237

243 DETLTLRGL 251

238 DSFTLTKGL 246

RESULT 3

AAU01243

ID AAU01243 standard; Protein; 258 AA.

XX AC AAU01243;

XX 18-JUL-2001 (first entry)

DE B. subtilis novel pantothenate kinase encoded by the gene coax.

Pantothenate kinase; coax; pantothenate biosynthesis; vitamin B5;

nutritional supplement; panto-compound; pantoate.

Bacillus subtilis.

WO200121772-A2.

29-MAR-2001.

21-SEP-2000; 2000WO-US25993.

21-SEP-1999; 99US-0400494.

07-JUN-2000; 2000US-0210072.

28-JUL-2000; 2000US-0221836.

24-AUG-2000; 2000US-0227860.

(OMNI-) OMNIGENE BIOPRODUCTS.

Yocum RR, Patterson TA, Hermann T, Pero JG;

DR WPI; 2001-218644/22.  
DR N-PSDB; AAS00984.  
XX  
PT New recombinant microorganism which overexpress a Bacillus subtilis  
PT pantothenate biosynthetic enzyme, useful for the high yield production  
PT of panto-compounds such as pantothenate and pantoate  
XX  
PS Example 14; Fig 23; 292pp; English.  
XX  
CC The sequence represents a novel B. subtilis pantothenate kinase (encoded  
CC by gene coax), an enzyme of the pantothenate biosynthetic pathway.  
CC Pantothenate, also known as vitamin B5, is used as a nutritional  
CC supplement in mammals and humans. The invention concerns methods of  
CC producing recombinant microorganisms overexpressing at least one Bacillus  
CC subtilis pantothenate biosynthetic enzyme. The microorganisms and methods  
CC of producing them are useful for producing a panto-compound such as  
CC pantothenate or pantoate, which is a nutritional requirement for  
CC livestock and humans. The methods are also useful for the identification  
CC of pantothenate kinase modulators. Panto-compounds are produced at a  
CC significantly higher yield than prior art methods and can be produced  
CC independent of the need to feed precursors which decreases expense.  
XX  
SQ Sequence 258 AA;

Query Match 32.3%; Score 428; DB 22; Length 258;  
Best Local Similarity 37.4%; Pred. No. 1.4e-36;  
Matches 98; Conservative 51; Mismatches 99; Indels 14; Gaps 6;  
QY 6 LLAVDIGNTTTVLGLADASGALHTWRTNREMLPDDDLALQLHGLFTLAGAPIPR--AA 63  
DB 2 LLVIDVGNNTVVLGVYH-DGKLEYHWRIETSRHKTEDEFGMILRSLSFDHSGLMFEQIDGI 60  
QY 64 VLSSVAPPVGENYALAKR----HFMDAFVSAENLPDVTVELDTPGSGVADRLCNLFG 119  
DB 61 IISVVVPPI----MFALERMCTKYFHIEPQIVGPGMKTGLNPKYDNPKEVGADRIVNAVA 116  
QY 120 AEKYLGGLDYAVVVDFTSTNFDVVGRRRFLGGILATGAQVSADALFARAACKLPRIITLQ 179  
DB 117 AIHLVG--NPLIVVDFTATTTCYIDENKQYMGGAIAPIGITISTEALYSRAAKLPRIET 174  
QY 180 APETAIGKNTVHALQSLGVFGYAEVMDGLLRIRRAELPGEAVAVATGGFSRTVQIGCQEI 239  
DB 175 RPDNIGKNTVSAMQSGILFGYVQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCI 234  
QY 240 DYYDETTLRLGLVELWASRSEV 261  
DB 235 DIVDPFLTLKGL-ELIYERNRV 255

RESULT 4  
AAU91149  
ID AAU91149 standard; Protein; 258 AA.  
XX  
AC AAU91149;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Bacillus subtilis pantothenate kinase Coax.  
XX  
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX  
OS Bacillus subtilis.  
XX  
PN WO200216601-A2.  
XX  
PD 28-FEB-2002.  
XX  
PF 24-AUG-2001; 2001WO-US26531.  
XX  
PR 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX

PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX  
PI Yocum RR, Patterson TA;  
XX  
DR WPI; 2002-269358/31.  
DR N-PSDB; ABK54168.  
XX  
PT Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein  
XX  
PS Claim 10; Page 67-68; 128pp; English.  
XX  
CC The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.  
XX  
SQ Sequence 258 AA;

Query Match 32.3%; Score 428; DB 23; Length 258;  
Best Local Similarity 37.4%; Pred. No. 1.4e-36;  
Matches 98; Conservative 51; Mismatches 99; Indels 14; Gaps 6;  
QY 6 LLAVDIGNTTTVLGLADASGALHTWRTNREMLPDDDLALQLHGLFTLAGAPIPR--AA 63  
DB 2 LLVIDVGNNTVVLGVYH-DGKLEYHWRIETSRHKTEDEFGMILRSLSFDHSGLMFEQIDGI 60  
QY 64 VLSSVAPPVGENYALAKR----HFMDAFVSAENLPDVTVELDTPGSGVADRLCNLFG 119  
DB 61 IISVVVPPI----MFALERMCTKYFHIEPQIVGPGMKTGLNPKYDNPKEVGADRIVNAVA 116  
QY 120 AEKYLGGLDYAVVVDFTSTNFDVVGRRRFLGGILATGAQVSADALFARAACKLPRIITLQ 179  
DB 117 AIHLVG--NPLIVVDFTATTTCYIDENKQYMGGAIAPIGITISTEALYSRAAKLPRIET 174  
QY 180 APETAIGKNTVHALQSLGVFGYAEVMDGLLRIRRAELPGEAVAVATGGFSRTVQIGCQEI 239  
DB 175 RPDNIGKNTVSAMQSGILFGYVQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCI 234  
QY 240 DYYDETTLRLGLVELWASRSEV 261  
DB 235 DIVDPFLTLKGL-ELIYERNRV 255

RESULT 5  
AAU91172  
ID AAU91172 standard; Protein; 258 AA.  
XX  
AC AAU91172;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Pantothenate kinase (Coax) #10.  
XX  
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX  
OS Bacillus stearothermophilus.  
XX  
PN WO200216601-A2.  
XX  
PD 28-FEB-2002.  
XX  
PF 24-AUG-2001; 2001WO-US26531.  
XX  
PR 24-AUG-2000; 2000US-227860P.  
PR

PR 20-MAR-2001; 2001US-0813453.  
XX  
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX  
PI Yocum RR, Patterson TA;  
XX  
DR WPI; 2002-269358/31.  
DR N-PSDB; ABK54193.

Identifying potential antibiotic or antimicrobial agent, comprises contacting composition comprising pantothenate kinase (Coax) protein with test compound and identifying inhibitor of the Coax protein -

Claim 10; Page 101-102; 128pp; English.

The invention describes assays for identifying a (potential) antibiotic comprising contacting an assay composition comprising a pantothenate kinase (Coax) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulating agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (Coax) protein described in the invention.

Sequence 258 AA;

Query Match 31.4%; Score 415.5; DB 23; Length 258; .  
 Best Local Similarity 35.2%; Pred. No. 2.9e-35;  
 Matches 94; Conservative 49; Mismatches 99; Indels 25; Gaps 6;

QY	6	LLAVDIGNTTTLVGLADASGALTHTWIRTNREMLPDDDLALQLHGLFTLAGAPIP	--RAA	63
Db	2	IFVLDVGNTNTVLGVYDGD--ELKHHWRIETSRKTEDEYGMKALLNHVGLQFSDIRGI	60	
QY	64	VLSSVAPPVGENYALAKR---HFMDAFVSAENLPDVTVELDTPGSGADRLCN---	116	
Db	61	IISSVVPPI---MFALERMLKYFHIKPLIVGPGIKTGLDIKYDNPREVGADRVNAVA	116	
QY	117	---LFGAEKYLGGLDYAVVVDFTSTNFDVWGRRRFLGGILATGAQVSADALFARAACL	173	
Db	117	GIHLYGSP-----LIIVDFGTATYCYINEHKQYMGGAIPGIMISTEALFARAACL	168	
QY	174	PRITLQAPETAIGKNTVHALQSLVFGYAEVMDGLLRIRRAELPGEAVAVATGGFSRTVQ	233	
Db	169	PRIEIRPDDIIGKNTVSAMQAGILYGVQVEGIVSRMKAKSKIPPVKVIATGGLAPLIA	228	
QY	234	GICQEIDYDETLTLRGLVELWASRSE	260	
QY	229	SESDIIDVVDPPFLTGLKLLYEKNT	255	

RESULT 6  
AAU911154  
ID AAU911154 standard; Protein; 255 AA.  
XX  
AC AAU911154;

Pantothenate kinase; CoaX; antibiotic; antimicrobial; pantothenate kinase modulator; coenzyme A; bactericidal compound.  
Geobacter sulfurreducens.

WO200216601-A2.

28-FEB-2002.

24-AUG-2001; 2001WO-US26531.

XX 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX  
XX  
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX  
XX Yocum RR, Patterson TA;  
PI  
XX  
XX WPI; 2002-269358/31.  
DR

Identifying potential antibiotic or antimicrobial agent, comprises contacting composition comprising pantothenate kinase (Coax) protein with test compound and identifying inhibitor of the Coax protein -

Claim 10; Page 72-73; 128pp; English.

The invention describes assays for identifying a (potential) antibiotic comprising contacting an assay composition comprising a pantothenate kinase (Coax) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulating agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (Coax) protein described in the invention.

Sequence 255 AA;

Query Match	31.3%;	Score 414.5;	DB 23;	Length 255;
Best Local Similarity	36.8%;	Pred. No. 3.7e-35;		
Matches 96;	Conservative 50;	Mismatches 102;	Indels 13;	Gaps 6

QY	6	LLAVDIGNTTTVLGLADASGALTHTWIRTNREMLPDDDLALQLHGLFTLAGAPIP--RAA	63
Dd	2	:             :      :    : :        :	
		2 LLVIDVGNINVLGIYDGE-RLVRDWRVSTDKARTTDEYGILLNELFRLAGLGLDQIRAV	60
QY	64	VLSSVAPP---VGENYALALKRHF MIDAFVSAENLPDVTVELDPGSVGADRLCN-LFG	119
Dd	61	:               :   :   :    :     :   :	
		61 IISSVPPLTGVLRLSLG--YFGMRPLVVGP GIKTMPLOYDNPREVGADRIVNAVAG	117
QY	120	AEKYLGGLDYAVVDFTSTNFDVWGRGRFLGGILATGAQVSADALFAARAACLPRITLQ	179
Dd	118	:       :      :    :    :     :     : :	
		118 YEKYRTSL---IIVDFGTATTDFDVNRKGEYCGGAIAPGLVISTEALFORASKLPRVDII	174
QY	180	APETAIGKNTVHALQSGLVFEGYAEMVDGLLRRIARAE L PGEAVAVATGGFSRTVQGICQEI	239
Dd	175	:     :    :    :    :     :     : :	
		175 RPSAIIARNTVNSMQAGIYYGVGLVDEIVTRMKAESKDAPRVIATGGLASLIAPEskTI	234
QY	240	DYDETTLTRLGLVELWASRSSE	260
Dd	235	:             :	
		235 EAVEEYLTLEGLRLIYNRRNE	255

RESULT 7  
AAU91151  
ID AAU91151 standard; Protein; 265 AA.  
XX  
AC AAU91151;

DE	<i>Streptomyces coelicolor</i> pantothenate kinase Coax.
XX	
KW	Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW	pantothenate kinase modulator; coenzyme A; bactericidal compound

WO200216601-A2.

PD 28-FEB-2002.



PF 24-AUG-2001; 2001WO-US26531.  
XX  
PR 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX  
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX  
PI Yocum RR, Patterson TA;  
XX  
DR WPI; 2002-269358/31.  
XX

PT Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -

PS Claim 10; Page 69-70; 128pp; English.

XX The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.

XX Sequence 265 AA;

Query Match 30.9%; Score 409.5; DB 23; Length 265;  
Best Local Similarity 37.9%; Pred. NO. 1.3e-34;  
Matches 97; Conservative 44; Mismatches 102; Indels 13; Gaps 4;

QY 6 LLAVDIGNTTTLGLADASGALHTWRTNREMLPDDLALQLHGLFTL-----AGAP 58  
DB 2 LLTIDVGNTHTVLGLDGEDIVEH-WRISTDSRRTADELAVLLQGLMGHPLLGLDELGDG 60  
QY 59 IPRAAVLSSVAPPVGENYALALKRHFMDAFVSAENLPDVTVELDTPGSGADRLCNLF 118  
DB 61 IDGIAICATVPSVLHELREVTTRYGDPVAVLVEPGVKTGPILTDPKKEVGADRIINAV 120  
QY 119 GAKEYLGLDYAVVVDFTSTNFDVVGRRRFLGGILATGAQVSADALFARAACKLPRLTL 178  
DB 121 AAVELYGG--PAIVVDFTATTFDAVSARGEYIGGVIAPIGIEISVEALGVKGAQLRKIEV 178  
QY 179 QAPETAIGKNTVHALQSLGVGYAEMVDGLLRIRRAEL--PGEAVAVATGGFSRTVQGI 235  
179 APRSRVIGKNTVEAMQSGIVYGFAGQVDGVNRMARELADDDPDDVTVIATGGGLAPMVLGE 238  
QY 236 QCEIDYDETLTLRGL 251  
DB 239 SSVIDEHEPWLTLMGL 254

RESULT 8  
AAU911170  
ID AAU911170 standard; Protein; 262 AA.  
XX  
AC AAU911170;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Pantothenate kinase (Coax) #8.

XX Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX  
OS Bacillus anthracis.  
XX  
PN WO200216601-A2.  
XX  
PD 28-FEB-2002.

XX 24-AUG-2001; 2001WO-US26531.  
XX  
PR 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX  
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX  
PI Yocum RR, Patterson TA;  
XX  
DR WPI; 2002-269358/31.  
DR N-PSDB; ABK54191.

XX Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -

PS Claim 8; Page 98-99; 128pp; English.

XX The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.

XX Sequence 262 AA;

Query Match 30.6%; Score 405.5; DB 23; Length 262;  
Best Local Similarity 31.7%; Pred. NO. 3.4e-34;  
Matches 84; Conservative 60; Mismatches 104; Indels 17; Gaps 4;

QY 6 LLAVDIGNTTTLGLADASGALHTWRTNREMLPDDLALQLHGLFTLAGAPIP--RAA 63  
DB 2 IFVLDVGNTHTVLGLGVFE-EGELRQHWRMETDRHKTDEYGMVLVKQLLEHGLSFEDVKGI 60  
QY 64 VLSSVAPPVGENYALALKRHFMDAFVSAENLPDVTVELDTPGSGVADRLCN-----L 117  
DB 61 IVSSVPPIMFALERMCCKYFKIKPLVVGPGIKTGLNIKYENPREVGADRIVNAVAGIHL 120  
QY 118 FGAKEYLGLDYAVVVDFTSTNFDVVGRRRFLGGILATGAQVSADALFARAACKLPRLT 177  
DB 121 YGSP-----LIIVDFGTATTTCYINEEKHYMGGVITPGIMISAEALYSRAAKLPRIE 172  
QY 178 LQAPETAIGKNTVHALQSLGVGYAEMVDGLLRIRRAELPGEAVAVATGGFSRTVQGIQ 237  
DB 173 ITPSSVVGNKNTVSAMQSGILYGVVGQVEGIVKRMKEAKQEPKVIATGGGLAKLISEEN 232  
QY 238 EIDYDETLTLRGLVELWASRSEVR 262  
DB 233 VIDVDPFLKGLKGLYMLYERNANLQ 257

RESULT 9  
AAG81225  
ID AAG81225 standard; Protein; 272 AA.  
XX  
AC AAG81225;  
XX  
DT 04-SEP-2001 (first entry)

XX Mycobacterium tuberculosis potential drug target protein SEQ ID 276.  
XX Drug target; growth; organism viability; characterisation.  
XX  
OS Mycobacterium tuberculosis.  
XX  
PN WO200135317-A1.  
XX

PD 17-MAY-2001.  
XX 13-NOV-2000; 2000WO-US31152.  
PF 12-NOV-1999; 99US-0165086.  
PR 12-NOV-1999; 99US-0165124.  
PR 01-FEB-2000; 2000US-0179531.  
XX (REGC ) UNIV CALIFORNIA.

XX Eisenberg D, Rotstein SH, Marcotte EM;

XX WPI; 2001-329193/34.  
DR N-PSDB; AAH52076.

XX Identifying nucleotide or polypeptide sequence for use as drug target,  
PT involves providing algorithm that analyzes a functional relationship  
PT between nucleotide or polypeptide sequences, and comparing the  
PT sequences

XX Disclosure; Page 188; 207pp; English.

CC This invention relates to a method for identifying a nucleotide or  
CC polypeptide sequence that may be a drug target, or essential for growth  
CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092  
CC represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium  
CC tuberculosis proteins which are potential drug targets. The DNA and  
CC protein sequences are used to illustrate the method of the invention. The  
CC method involves providing an unknown nucleotide or polypeptide sequences,  
CC and comparing it to a number of sequences along with at least one  
CC algorithm capable of analysing a functional relationship between  
CC nucleotide and polypeptide sequences. The method is useful for  
CC characterising the function of nucleic acids and polypeptides that may be  
CC useful as a target for a drug or essential for the growth or viability of  
CC an organism.

XX Sequence 272 AA;

Query Match 30.5%; Score 404; DB 22; Length 272;

Best Local Similarity 38.3%; Pred. No. 5.1e-34;

Matches 106; Conservative 42; Mismatches 95; Indels 34; Gaps 7;

QY 6 LLAVDIGNTTVLGLADASGALHTHT-----WRIRTNREMLPDDDLALQLHGLFTLAGAPIP 60  
Db 2 LLAIDVRNTHTVVGL--LSGMKEHAKVQWRIQRTSEVTADLALTDGLIGEDSERLT 59  
QY 61 RAAVLSSVAPPVGENYALAKRHFMDAFVSAENLPDVTVE-----LDTPGSVG 110  
60 GTAALSTV-PSVLHEVRIMLDQYW-----PSVPHVLIIEPGVRTGIPLLVDPNKEVG 109  
111 ADRLCNLFGEAEKYLGLDYAVVVDFTSTNFDVVGRRRFLGGILATGAQVSADALFARA 170  
Db 110 ADRIVNCCLAA--YDRFRKAAIVVDFGSSICVDVVSAGKEFLGGAIAPGVQVSSDAAAARS 167  
QY 171 AKLPRITLQAPETAIGKNTVHALQSLVFGYAEWVDGLLRIRRAELPG-----EAVAVAT 225  
Db 168 AALRRVELARPRSVGKNTVECMQAGAVFGFAGLVGVRIGREDVSGFSDHDAIVAT 227  
QY 226 GGFSRTVQICQEIYDETLTLRGLVVELWASRSEVR 262  
Db 228 GHTAPLLPELHTVDHYDQHLTLQGLRLVLFERNLEQV 264

RESULT 10

AAU911152

ID AAU911152 standard; Protein; 272 AA.

XX AAU911152;

XX 05-JUN-2002 (first entry)

XX Mycobacterium tuberculosis pantothenate kinase Coax.

KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX Mycobacterium tuberculosis.  
OS WO200216601-A2.  
PN 28-FEB-2002.  
XX 24-AUG-2001; 2001WO-US26531.  
PF 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
PA Yocum RR, Patterson TA;  
XX WPI; 2002-269358/31.

XX Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein .

XX Claim 10; Page 70-71; 128pp; English.

XX The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.

XX Sequence 272 AA;

Query Match 30.5%; Score 404; DB 23; Length 272;

Best Local Similarity 38.3%; Pred. No. 5.1e-34;

Matches 106; Conservative 42; Mismatches 95; Indels 34; Gaps 7;

QY 6 LLAVDIGNTTVLGLADASGALHTHT-----WRIRTNREMLPDDDLALQLHGLFTLAGAPIP 60  
Db 2 LLAIDVRNTHTVVGL--LSGMKEHAKVQWRIQRTSEVTADLALTDGLIGEDSERLT 59  
QY 61 RAAVLSSVAPPVGENYALAKRHFMDAFVSAENLPDVTVE-----LDTPGSVG 110  
Db 60 GTAALSTV-PSVLHEVRIMLDQYW-----PSVPHVLIIEPGVRTGIPLLVDPNKEVG 109  
111 ADRLCNLFGEAEKYLGLDYAVVVDFTSTNFDVVGRRRFLGGILATGAQVSADALFARA 170  
Db 110 ADRIVNCCLAA--YDRFRKAAIVVDFGSSICVDVVSAGKEFLGGAIAPGVQVSSDAAAARS 167  
QY 171 AKLPRITLQAPETAIGKNTVHALQSLVFGYAEWVDGLLRIRRAELPG-----EAVAVAT 225  
Db 168 AALRRVELARPRSVGKNTVECMQAGAVFGFAGLVGVRIGREDVSGFSDHDAIVAT 227  
QY 226 GGFSRTVQICQEIYDETLTLRGLVVELWASRSEVR 262  
Db 228 GHTAPLLPELHTVDHYDQHLTLQGLRLVLFERNLEQV 264

RESULT 11

AAU911173

ID AAU911173 standard; Protein; 260 AA.

XX AAU911173;

XX 05-JUN-2002 (first entry)

XX Pantothenate kinase (Coax) #11.

[illegible]

RESULT 12  
AAU91163  
ID AAU91163 standard; Protein; 233 AA.  
XX  
XX AAU91163;  
XX  
DT 05-JUN-2002 (first entry)

XX	Pantothenate kinase (Coax) #1.
DE	Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX	pantothenate kinase modulator; coenzyme A; bactericidal compound.
KW	Bacillus subtilis.
OS	WO200216601-A2.
XX	28-FEB-2002.
PN	24-AUG-2001; 2001WO-US26531.
PD	24-AUG-2000; 2000US-227860P.
XX	20-MAR-2001; 2001US-0813453.
XX	(OMNI-) OMNIGENE BIOPRODUCTS INC.
PA	Yocum RR, Patterson TA;
XX	WPI; 2002-269358/31.
DR	N-PSDB; ABK54169.
XX	Identifying potential antibiotic or antimicrobial agent, comprises
PT	contacting composition comprising pantothenate kinase (Coax) protein
PT	with test compound and identifying inhibitor of the Coax protein -
PT	Disclosure; Page 81-82; 128pp; English.
PS	The invention describes assays for identifying a (potential) antibiotic
XX	comprising contacting an assay composition comprising a pantothenate
CC	kinase (Coax) protein with a test compound, and determining the ability
CC	of the test compound to inhibit the activity of the Coax protein, an
CC	essential enzyme for the production of coenzyme A. Coax protein is a
CC	valuable target for identifying bactericidal compounds. Coax modulating
CC	agents can be used in an infectious animal model to determine the
CC	efficacy, toxicity, or side effects of treatment with such an agent. This
CC	is the amino acid sequence of a pantothenate kinase (Coax) protein
CC	described in the invention.
XX	Sequence 233 AA;
SQ	Query Match 28.6%; Score 378.5; DB 23; Length 233;
	Best Local Similarity 38.3%; Pred. No. 2e-31;
	Matches 82; Conservative 44; Mismatches 75; Indels 13; Gaps 5
QY	6 LLAVDIGNTTTLVGLADASGALTHTWRITNRMLPDDLLQLHLFTLAGAPIPR--AA 63
Db	: :        :     :      :  :      :  :  :  :  :
Db	2 LLVIDVGNTNTVLGVYH-DGKLEYHWRIETSRHKTEDEFGMILRSLSFDHSGLMFEQIDGI 60
QY	64 VLSSVAPPVGENYALAKR---HFMDAFAVSAENLPDVTVELDTPGSVGADRLCNLFG 119
Db	:      :     :  :  :  :  :  :  :  :  :  :  :  :  :
Db	61 IISVVPPI----MFALERMTKYFHIEPQIVPGMKTGKLNKIKYDNPKVEGDRIVNVA 116
QY	120 AEKYLGGLDYAVVVDGETSTNFVDVGRRRFLGGILATGAQVSADALFARAACLPRITLQ 179
Db	:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db	117 AIHLYG--NPLIVVDFGTATTTCYIDENKQYMGGAIAPGITISTEALYSRAKLPRIEIT 174
QY	180 APETAIGKNTVHALQSGLVFGEYAEMVDGLLRIR 213
Db	:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db	175 RPDNIIGKNTVSAMOSGILFGYVGQVEGIVKRMK 208

RESULT 13	
AAU91175	
ID	AAU91175 standard; Protein; 256 AA.
XX	
AC	AAU91175;
XX	
DT	05-JUN-2002 (first entry)
XX	
DE	Pantothenate kinase (Coax) #13.

XX Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX Clostridium difficile.  
OS WO200216601-A2.  
XX 28-FEB-2002.  
PN 24-AUG-2001; 2001WO-US26531.  
PD 24-AUG-2000; 2000US-227860P.  
XX 20-MAR-2001; 2001US-0813453.  
PR (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX Yocum RR, Patterson TA;  
PI WPI; 2002-269358/31.  
XX N-PSDB; ABK54196.  
PT Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -  
XX Claim 6; Page 105; 128pp; English.  
PS The invention describes assays for identifying a (potential) antibiotic  
XX comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
XX described in the invention.  
SQ Sequence 256 AA;  
Query Match 28.5%; Score 377; DB 23; Length 256;  
Best Local Similarity 35.7%; Pred. No. 3.3e-31;  
Matches 94; Conservative 51; Mismatches 102; Indels 16; Gaps 9;  
QY 6 LLAVDIGNTTTVGLADASGALHTWRTNRNREMLPDDALQLHGLFTLAGAPIRA--A 63  
Db 2 LLVFDVGNNTNMVLGIYKGD-KLVNYWRIKTDREKTSDEYGILISNLFYDYNVISDIDDV 60  
Db 64 VLSSVAPPVG---ENYALA-LKRHFMDAFVSAENLPDVTVELDTPGSGADRLCN-LF 118  
Db 61 IISVVVPMVHNSLENFCIKYCKKQPLIVGPGIKT----GLNIXYDNPKQVGADRIVNAVA 116  
QY 119 GAELYGLLDYAVVDFGTSTNFDVVGRRRFLGGILATGAQVSADALFARAAKLPRITL 178  
Db 117 GIEKY--GAP-SILVDFGTATTFCATSEKGEYLGGTIAPGKISSEALFQSASKLPRVEL 173  
QY 179 QAPETAIGKNTVHALQSLVFGYAEVMVDGLLRIRAEPLGEAV-AVATGGFSRTVQIGICQ 237  
Db 174 AKPGMTICKSTVSAMQSGIIVGYGLVDKIIISIMKKELNCDDVKVIATGGLAKLIASETK 233  
QY 238 EIDYVDETLTLRGLVELWASRSE 260  
Db 234 SIDYVDGFLTLEGLRIIYEKNQE 256  
RESULT 14  
ABB47661  
ID ABB47661 standard; Protein; 259 AA.  
XX ABB47661;  
AC ABB47661;  
XX 05-FEB-2002 (first entry)

XX Listeria monocytogenes protein #365.  
DE Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
XX vitamin B12; bacterial infection; disease.  
KW Listeria monocytogenes.  
XX WO200177335-A2.  
XX 18-OCT-2001.  
XX 11-APR-2001; 2001WO-FR01118.  
PF 11-APR-2000; 2000FR-0004629.  
XX (INSP ) INST PASTEUR.  
PA Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;  
XX Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;  
PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;  
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;  
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;  
PI Rose M, Voss H;  
XX WPI; 2002-010914/01.  
DR Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
XX and prevention of Listeria and related bacterial infections, and  
PT related polypeptides -  
PT Claim 6; SEQ ID No 366; 192pp; French.  
XX The present invention relates to the genome sequence of Listeria  
CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of  
CC it are useful for selecting probes and primers for detecting genes in L.  
CC monocytogenes and related organisms, and for studying genetic  
CC polymorphisms and other genomes. The present sequence is a protein  
CC encoded by the genome sequence of the present invention. Proteins  
CC expressed from the genome sequence are useful for raising specific  
CC antibodies, identification of L. monocytogenes and related organisms, and  
CC for biosynthesis and biodegradation, especially biosynthesis of vitamin  
CC B12. The genome sequence and proteins encoded by it are also useful for  
CC selecting compounds that regulate gene expression and cell replication  
CC and modulate L. monocytogenes-related diseases. In addition, the genome  
CC sequence and proteins encoded by it are useful in pharmaceutical and  
CC vaccines compositions for the treatment or prevention of infections by L.  
CC monocytogenes and related organisms.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX Sequence 259 AA;  
SQ Query Match 27.5%; Score 363.5; DB 23; Length 259;  
Best Local Similarity 31.7%; Pred. No. 8.8e-30;  
Matches 80; Conservative 53; Mismatches 114; Indels 5; Gaps 4;  
QY 6 LLAVDIGNTTTVGLADASGALHTWRTNRNREMLPDDALQLHGLFTLAG-APIP-RAA 63  
Db 2 ILVIDVGNTNCTGVYEKQKLLKH-WRMTTDRHRTSDELGMTVLNFFSYANLTPSDIQGI 60  
QY 64 VLSSVAPPVGENYALAKRHFMDAFVSAENLPDVTVELDTPGSGADRLCNLFGAEKY 123  
Db 61 IISVVVPPIMHAMETMVCVRYFNIRPLIVGPGTKTGLNKLKVDNPREIGSDRIVNVAASEE 120  
QY 124 LGGLDYAVVVDFGTSTNFDVVGRRRFLGGILATGAQVSADALFARAAKLPRITLOAPET 183  
Db 121 YG--TPVIVVDFGTATTFCYIDESGVYQGGALAPGIMISTEALYNRAAKLPRVDIAESSQ 178  
QY 184 AIGKNTVHALQSLVFGYAEVMVDGLLRIRAEPLGEAVAVATGGFSRTVQIGCQEIDYD 243



Db 179 IIGKSTVSSMQAGIFYGFVGQCEGIIAEMRKQSNASPVVATGGGLARMITEKSSAVDILD 238  
QY 244 ETLTLRGLVELW 255  
Db 239 PFLTLKGLELLY 250

RESULT 15

AAU91153  
ID AAU91153 standard; Protein; 258 AA.

XX AC AAU91153;

XX DT 05-JUN-2002 (first entry)

XX DE Rhodobacter capsulatus pantothenate kinase Coax.

XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
pantothenate kinase modulator; coenzyme A; bactericidal compound.

OS Rhodobacter capsulatus.

PN WO200216601-A2.

XX PD 28-FEB-2002.

XX PF 24-AUG-2001; 2001WO-US26531.

XX PR 24-AUG-2000; 2000US-227860P.

XX PR 20-MAR-2001; 2001US-0813453.

XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX PI Yocum RR, Patterson TA;

XX DR WPI; 2002-269358/31.

XX PT Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -

PS Claim 10; Page 71-72; 128pp; English.

XX CC The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.

XX SQ Sequence 258 AA;

Query Match 26.2%; Score 347.5; DB 23; Length 258;  
Best Local Similarity 36.0%; Pred. No. 4.3e-28;  
Matches 90; Conservative 42; Mismatches 109; Indels 9; Gaps 6;

QY 6 LLAVDIGNTTTVLGLADASCALTHTWIRTNREMLPDDLALQLHGLFTLAGAP-IPRAAV 64  
Db 2 LLCIDCGTNTVFSVWDGTD-FAATWRIATDHRRTADEYFVWLNTLMQLKLGQRISAI 60

QY 65 LSSVAPPVGENYALAKRHEMIDAFVSAE--NLPDVTVELDTPGSVGADRLCN-LFGAE 121  
Db 61 ISSAPRVWFNLRLVLCNRYDCRPYVVGKGPCELP-VAPRPDPGTTVGPDLVNTVAGYD 119

QY 122 KYLGGLDYAVVWDFGTSTNFDVVGRRRFLGGILATGAQVSADALFARAAKLPRITLQAP 181  
Db 120 RHGGDL---IVVDFGTATFDVVPADGAYIGGVIAPGVNLNLSLEALHMAAALPHVDYTKP 176

QY 182 ETAIGKNTVHALQSLVFGYAEMVDGLLRIRRAELPGEAVAVATGGFSRTVQGICQEIDY 241  
Db 177 QGVIGTNTVACIQSGVTWGYIGLVEGIVRQIRMERDRPMKVIATGGLASLFDLGFDFDK 236  
QY 242 YDETTLTLRGL 251  
Db 237 VEDDLTMHGL 246

Search completed: June 24, 2003, 21:46:16  
Job time : 27.899 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:36:04 ; Search time 9.55623 Seconds  
(without alignments)  
2635.685 Million cell updates/sec

Title: US-09-813-453A-8  
Perfect score: 1324  
Sequence: 1 MPAPFLLAVDIGNTTVLGL.....DETTLRLGLVELWASRSEVR 262

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

--Searched: 283224 seqs, 96134422 residues  
otal number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1324	100.0	262	2 E75516	conserved hypothetical
2	437.5	33.0	254	2 F83660	hypothetical prote
3	409.5	30.9	265	2 T36391	hypothetical prote
4	404	30.5	272	2 A70955	hypothetical prote
5	401	30.3	274	2 H86937	conserved hypothet
6	396.5	29.9	261	2 B87489	transcription acti
7	378.5	28.6	233	2 S66100	conserved hypothet
8	371.5	28.1	273	2 E97293	probable transcrip
9	363.5	27.5	259	2 AF1102	conserved hypothet
10	356.5	26.9	259	2 AF1464	conserved hypothet
11	260	19.6	246	2 D72320	conserved hypothet
12	248	18.7	273	2 D71326	conserved hypothet
13	159	12.0	262	2 F70165	conserved hypothet
14	152.5	11.5	276	2 AI2292	hypothetical prote
15	133.5	10.1	229	2 E70465	hypothetical prote
16	128.5	9.7	242	2 A82637	conserved hypothet
17	123.5	9.3	257	2 S75559	hypothetical prote
18	104.5	7.9	400	2 T29121	hypothetical prote
19	103.5	7.8	267	2 I40327	baf protein - Bord
20	102.5	7.7	675	2 A75627	H+/K+-exchanging A
21	101.5	7.7	444	2 D87557	major facilitator
22	101	7.6	636	2 T35182	probable ABC-type
23	101	7.6	1441	2 B86807	hypothetical prote
24	100.5	7.6	592	2 H82031	probable biotin-[a
25	99.5	7.5	544	2 H72647	hypothetical prote
26	98.5	7.4	302	2 G70614	hypothetical prote
27	96.5	7.3	515	2 AF3524	hypothetical prote
28	96.5	7.3	592	2 B81009	BirA protein/Bvg a
29	96	7.3	360	2 S06287	fragmin - slime mo

30	94.5	7.1	314	2 H97764	malate dehydrogena
31	94	7.1	485	2 F87448	succinylglutamic s
32	93.5	7.1	419	2 D69202	5-enolpyruvylshik
33	91.5	6.9	998	2 G83022	probable two-compo
34	91	6.9	439	2 F84278	3-phosphoshikimate
35	90.5	6.8	258	1 G69985	probable enoyl-CoA
36	90.5	6.8	317	2 F72246	glucokinase - Ther
37	90.5	6.8	346	2 T31263	xylene monooxygena
38	90.5	6.8	628	2 AC2261	cell division prot
39	90	6.8	519	2 AI0400	glutamate-cysteine
40	89.5	6.8	332	2 A86892	glycosyltransferas
41	89.5	6.8	405	2 H81995	probable transmemb
42	89.5	6.8	499	2 E82977	glycolate oxidase
43	89.5	6.8	503	2 T37119	probable membrane-
44	89.5	6.8	530	2 AF2175	hypothetical prote
45	89.5	6.8	876	2 D70971	hypothetical prote

ALIGNMENTS

RESULT 1  
E75516

conserved hypothetical protein - Deinococcus radiodurans (strain R1)

C;Species: Deinococcus radiodurans

C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 20-Jun-2000

C;Accession: E75516

R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1

A;Reference number: A75250; MUID:20036896; PMID:10567266

A;Accession: E75516

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-262 <WHI>

A;Cross-references: GB:AE001905; GB:AE000513; NID:g6458144; PIDN:AAF10040.1; PID:g64

A;Experimental source: strain R1

C;Genetics:

A;Gene: DR0461

A;Map position: 1

C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 100.0%; Score 1324; DB 2; Length 262;  
Best Local Similarity 100.0%; Pred. No. 9.1e-101;  
Matches 262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MPAPFLLAVDIGNTTVLGLADASGALHTWIRTNREMLPDDDLALQLHGLFTLAGAPIP 60
Db	1	MPAPFLLAVDIGNTTVLGLADASGALHTWIRTNREMLPDDDLALQLHGLFTLAGAPIP 60
Qy	61	RAAVLSSVAPPVGENYALALKRHFMDAFVSAENLPDVTVELDTPGSGADRLCNLFGA 120
Db	51	RAAVLSSVAPPVGENYALALKRHFMDAFVSAENLPDVTVELDTPGSGADRLCNLFGA 120
Qy	121	EKYLGLDYAVVVDFTSTNFDVVGRRRFLGGILATGAQVSADALFARAACLPRITLQA 180
Db	121	EKYLGLDYAVVVDFTSTNFDVVGRRRFLGGILATGAQVSADALFARAACLPRITLQA 180
Qy	181	PETAICKNTVHALQSLVFGYAEMVDGLLRIRRAELPGEAVAVATGGFSRTVQGICQEID 240
Db	181	PETAICKNTVHALQSLVFGYAEMVDGLLRIRRAELPGEAVAVATGGFSRTVQGICQEID 240
Qy	241	YYDETTLRLGLVELWASRSEVR 262
Db	241	YYDETTLRLGLVELWASRSEVR 262

RESULT 2

F83660  
hypothetical protein BH0086 [imported] - Bacillus halodurans (strain C-125)  
C;Species: Bacillus halodurans







probable transcription regulator, homolog of Bvg accessory factor [imported] - Clostridi  
C;Species: Clostridium acetobutylicum  
C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001  
C;Accession: E97293  
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo  
A;Reference number: A96900; MUID:21359325; PMID:21359325  
A;Accession: E97293  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-273 <KUR>  
A;Cross-references: GB:AE001437; PIDN:AAK81136.1; PID:g15026270; GSPDB:GN00168  
A;Experimental source: Clostridium acetobutylicum ATCC824  
C;Genetics:  
A;Gene: CAC3200  
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 28.1%; Score 371.5; DB 2; Length 273;  
Best Local Similarity 34.5%; Pred. No. 5.8e-23;  
Matches 89; Conservative 51; Mismatches 97; Indels 21; Gaps 6;

QY 6 LLAVDIGNTTTLGLADASGALHTHTWRIRTNREMLPDDLALQLHGLF-----TLGAP 58  
Db 13 ILVLDVGNNTNVLGIYNDT-KLTAEWRLSTDLRSADYGIQVMNLFQQDKLDPTLV--- 68

QY 59 IPRAAVLSSVAPPVGENYALAKRHFMDAFVSAENLPDVTVELDTPGSGVADRILCNLF 118  
Db 69 --EGVISSVVPNIMYSLEHMIRKYFKINPLVVGPGIKTGINIKYDNPKEVGADRVNAV 126

QY 119 GA-EKYLGLDYAVVDFGTSTNFDVVGRRRFLGGILATGAQVSADALFARAACLPRIT 177  
Db 127 AAHEIYKRSL---IIIDFGTATTFCVAVRENGDYLGAICPGIKVSSSEALFEKAACLPRVE 183

QY 178 LQAPETAIGKNTVHALQSLVFGYAEWVDGLLRIRAEPLGEA---VAVATGGFSRTVQ 233  
Db 184 LKPAYAICKNTISSIQSGIVGYIGQVRYIVERMKEELOEKEPLVVATGGLAKLIS 243

QY 234 GICQEIYDYDETLRLGL 251  
Db 244 EEAKNVVDVINPFLLEGL 261

RESULT 9  
AF1102  
conserved hypothetical protein lmo0221 [imported] - Listeria monocytogenes (strain EGD-e  
C;Species: Listeria monocytogenes  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
Accession: AF1102  
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
A;Title: Comparative genomics of Listeria species.  
A;Reference number: AB1077; MUID:21537279; PMID:11679669  
A;Accession: AF1102  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-259 <GLA>  
A;Cross-references: GB:NC\_003210; PIDN:CAD00748.1; PID:g16409586; GSPDB:GN00177  
A;Experimental source: strain EGD-e  
C;Genetics:  
A;Gene: lmo0221  
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 27.5%; Score 363.5; DB 2; Length 259;  
Best Local Similarity 31.7%; Pred. No. 2.4e-22;  
Matches 80; Conservative 53; Mismatches 114; Indels 5; Gaps 4;

QY 6 LLAVDIGNTTTLGLADASGALHTHTWRIRTNREMLPDDLALQLHGLFTLAG-APIP-RAA 63

Db 2 ILVIDVGNTNCTVGVEYKQKLLKH-WRMTTDRHRTSDELGMTVLNFFSYANLTPSDIQGI 60

QY 64 VLSSVAPPVGENYALAKRHFMDAFVSAENLPDVTVELDTPGSGVADRILCNLFGAEKY 123  
Db 61 IISVVPPIMHAMETMCVRYFNIRPLIVPGIKTGLNLKVDNPREIGSDRIVNAVAASEE 120

QY 124 LGGLDYAVVDFGTSTNFDVVGRRRFLGGILATGAQVSADALFARAACLPRITLQAPET 183  
Db 121 YG--TPVIVVDFGTATTCYIDESGVYQGGAIAPGIMISTEALYNRAAKLPRVDIAESSQ 178

QY 184 AIGKNTVHALQSLVFGYAEWVDGLLRIRAEPLGEAVAVATGGFSRTVQICQEIYYD 243  
Db 179 IIGKSTVSSMQAGIFYGVGQCEGIIAEMKKQSNASPVVATGGLARMITEKSSAVDILD 238

QY 244 ETTLRLGLVELW 255  
Db 239 PFLTGLGLELLY 250

RESULT 10  
AF1464  
conserved hypothetical protein lin0253 [imported] - Listeria innocua (strain Clip112  
C;Species: Listeria innocua  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C;Accession: AF1464  
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloe  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.  
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehl  
A;Title: Comparative genomics of Listeria species.  
A;Reference number: AB1077; MUID:21537279; PMID:11679669  
A;Accession: AF1464  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-259 <GLA>  
A;Cross-references: GB:AL592022; PIDN:CAC95486.1; PID:g16412682; GSPDB:GN00178  
A;Experimental source: strain Clip11262  
C;Genetics:  
A;Gene: lin0253  
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 26.9%; Score 356.5; DB 2; Length 259;  
Best Local Similarity 31.7%; Pred. No. 9.1e-22;  
Matches 80; Conservative 52; Mismatches 115; Indels 5; Gaps 4;

QY 6 LLAVDIGNTTTLGLADASGALHTHTWRIRTNREMLPDDLALQLHGLFTLAG-APIP-RAA 63  
Db 2 ILVIDVGNTNCTVGVEYKQKLLRH-WRMTTDRHRTSDELGMTVLNFFSYANLTPSDIQGI 60

QY 64 VLSSVAPPVGENYALAKRHFMDAFVSAENLPDVTVELDTPGSGVADRILCNLFGAEKY 123  
Db 61 IISVVPPIMHAMETMCVRYFNIRPLIVPGIKTGLNLKVDNPREIGSDRIVNAVAASEE 120

QY 124 LGGLDYAVVDFGTSTNFDVVGRRRFLGGILATGAQVSADALFARAACLPRITLQAPET 183  
Db 121 YG--TPVIVVDFGTATTCYIDEGVYQGGAIAPGIMISTEALYNRAAKLPRVDIAESSQ 178

QY 184 AIGKNTVHALQSLVFGYAEWVDGLLRIRAEPLGEAVAVATGGFSRTVQICQEIYYD 243  
Db 179 IIGKSTVSSMQAGIFYGVGQCEGIIAEMKKQSNATSPVVVATGGLARMITEKSSAVDILD 238

QY 244 ETTLRLGLVELW 255  
Db 239 PFLTGLGLELLY 250

RESULT 11  
D72320  
conserved hypothetical protein - Thermotoga maritima (strain MSB8)  
C;Species: Thermotoga maritima







```
US-09-813-453A-5
; Sequence 5, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; -09-813-453A-5

Query Match      29.4%; Score 382.5; DB 9; Length 272;
Best Local Similarity 34.0%; Pred. No. 2.1e-28;
Matches 90; Conservative 53; Mismatches 107; Indels 15; Gaps 5;

QY      1 MLLAIEQGNNTMTFAIHDG----ASWVAQWRSATESRTRADEYVVVWLSQQLSMQGLGFRA 56
      |||||: |||: : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1 MLLAIDVRNTHTVVGLSGMKEHAKVQWQWRIRTESEVTADELALTIDGLI---GEDSER 57
      |||||: |||: : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      57 IDAVIISSVVPQSIFNLRNLSRRYF-NVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAI 115
      : | | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db      58 LTGTAALSTVPSVLHEVRIMLDQYWPSVPHVLIIEPGVRTGIPLLVNDPNKEVGADRIVNCL 117
      : | | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |

QY      116 GAAMVYGPPLVVIDSGTATTFDIIAADGAFEGGIIAPGINLSMQALHEAAAKLPRIAQR 175
      | : : | | : : | | | : : | : : | : : | : : | : : | : : | : : | :
Db      118 AAYDRFRKAAIVVDFGSSICVDVVSAGKEFLGGAIAAGVQVSSDAAARSAAALRRVELAR 177
      : : | | | : : | | | : : | : : | : : | : : | : : | : : | : : | :

QY      176 PAGNRIVGTDTVSAMQSGVFWGYISLIEGLVARIKAERG-----EPMTVIATGGVASLFE 230
      | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      178 P--RSVVGKNTVECMQAGAVFGFAGLV DGLVGRIREDVSGFSVDHDAIVATGHTAPLLL 235
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      231 GATDSIDHSDLTIRGLLEIYRN 255
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      236 PELHTVDHYDQHLTLQGLRLVFERN 260
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

.RESULT 15
3-09-712-363-276
; Sequence 276, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206,
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
```

```
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 276
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; US-09-712-363-276

Query Match      29.1%; Score 378.5; DB 9; Length 272;
Best Local Similarity 33.6%; Pred. No. 5.1e-28;
Matches 89; Conservative 54; Mismatches 107; Indels 15; Gaps 5;

QY      1 MLLAIEQGNNTMTFAIHDG----ASWVAQWRSATESRTRADEYVVVWLSQQLSMQGLGFRA 56
      : |||||: |||: : | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1 VLLAIDVRNTHTVVGLSGMKEHAKVQWQWRIRTESEVTADELALTIDGLI---GEDSER 57
      : |||||: |||: : | | | | | | | | | | | | | | | | | | | | | | | | |

QY      57 IDAVIISSVVPQSIFNLRNLSRRYF-NVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAI 115
      : | | | : : | : : | : : | : : | : : | : : | : : | : : | : : | :
Db      58 LTGTAALSTVPSVLHEVRIMLDQYWPSVPHVLIIEPGVRTGIPLLVNDPNKEVGADRIVNCL 117
      : | | | : : | : : | : : | : : | : : | : : | : : | : : | : : | :

QY      116 GAAMVYGPPLVVIDSGTATTFDIIAADGAFEGGIIAPGINLSMQALHEAAAKLPRIAQR 175
      | : : | | : : | | | : : | : : | : : | : : | : : | : : | : : | :
Db      118 AAYDRFRKAAIVVDFGSSICVDVVSAGKEFLGGAIAAGVQVSSDAAARSAAALRRVELAR 177
      : : | | | : : | | | : : | : : | : : | : : | : : | : : | : : | :

QY      176 PAGNRIVGTDTVSAMQSGVFWGYISLIEGLVARIKAERG-----EPMTVIATGGVASLFE 230
      | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      178 P--RSVVGKNTVECMQAGAVFGFAGLV DGLVGRIREDVSGFSVDHDAIVATGHTAPLLL 235
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      231 GATDSIDHSDLTIRGLLEIYRN 255
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      236 PELHTVDHYDQHLTLQGLRLVFERN 260
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Search completed: June 24, 2003, 22:29:03
Job time : 15.4494 secs
```



Qy 181 IVGTDVTSAMQSGVFWGYISLIEGLVARIKAEGERPMTVIATGCVASLFEATDSIDHFD 240  
Db 179 IIRNVTNSMQAGIYYGYVGLVDEIVTRMKAESKADPRVIATGGLASLIAPESKTIEAVE 238  
Qy 241 SDLTIRGLLEIYRN 255  
Db 239 EYLTLEGLRILYERN 253

RESULT 8  
US-09-813-453A-55  
; Sequence 55, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 55  
; LENGTH: 256  
; TYPE: PRT  
; ORGANISM: Clostridium difficile  
US-09-813-453A-55

Query Match 40.3%; Score 523.5; DB 9; Length 256;  
Best Local Similarity 40.6%; Pred. No. 8.6e-42;  
Matches 104; Conservative 53; Mismatches 96; Indels 3; Gaps 2;  
Qy 1 MLLAIEQGNNTMTMFAIHDGASWVAQWRSATESSTRTADEYVWVLSQLLSMQGLGFRDAIDAV 60  
Db 1 MLLVDVGNNTNMVLGIYKDKLVNFWRIKTDREKTSDEYGILISNLFDDYDNVNISDIDDV 60  
Qy 61 IISVVVPQSIFNLNLSRRYFNVEPLVIGENAKGLIDVRIEKPSEAGADRLVNAIGAAMV 120  
Db 61 IISVVPNVMHSLNFCIKYCKQPLIVGPGIKTGLNIKYNPNKQVGADRVNAVAGIEK 120  
Qy 121 YPGPLVIDSGTATTFDIAADGAFEGGIIAPGINLSMQALHEAAKLPRIAIQRPAGNR 180  
Db 121 YGAPSLVDFGTATTFCAISEKGEYLGTTIAPGIKISSEALFQASKLPVELAKP--GM 178  
Y 181 IVGTDVTSAMQSGVFWGYISLIEGLVARIKAEGERPMTVIATGCVASLFEATDSIDHF 239  
Db 179 TICKSTVSAMQSGIYYGYVGLVDKIISIMKELNCDVVKVIATGGLAKLIASETKSIDIV 238  
Qy 240 DSDLTIRGLLEIYRN 255  
Db 239 DGFLEGLRIIYEKN 254

RESULT 9  
US-09-813-453A-4  
; Sequence 4, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569

; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 265  
; TYPE: PRT  
; ORGANISM: Streptomyces coelicolor  
US-09-813-453A-4

Query Match 39.8%; Score 517; DB 9; Length 265;  
Best Local Similarity 42.4%; Pred. No. 3.7e-41;  
Matches 112; Conservative 50; Mismatches 90; Indels 12; Gaps 5;  
Qy 1 MLLAIEQGNNTMTMFAIHDGASWVAQWRSATESSTRTADEYVWVLSQLLSM-----QGLGFR 55  
Db 1 MLLTIDVGNTHVTLGLDFGDEDIVHWRISTDSRRTADELAVLQGLMGHPLLGLDELG-D 59  
Qy 56 AIDAVIISVVVPQSIFNLNLSRRYFNVEPLVIGENAKGLIDVRIEKPSEAGADRLVNA 114  
Db 60 GIDGIAICATVPSVLHELREVTTRYGDPVAVLVEPGVKTGPILTDDHPKEVGADRIINA 119  
Qy 115 IGAAVMYPGPLVVIDSATTFDIAADGAFEGGIIAPGINLSMQALHEAAKLPRIAIQ 174  
Db 120 VAAVELYGGPAIVDFGTATTFDAVSARGEYIGGVIAPGIEISVEALGVKGAQLRKIEVA 179  
Qy 175 RPAGNRIVGTDVTSAMQSGVFWGYISLIEGLV---ARIKAERGEPMTVIATGGVASLEEG 231  
Db 180 RP--RSVIGKNTVEAMQSGIVYGFAGQVGDGVVNRMARLADDDVTVIATGGLAPMVLG 237  
Qy 232 ATDSIDHFDSDLTIRGLLEIYRN 255  
Db 238 ESSVIDEHEPWLTLMLGLRLVYERN 261

RESULT 10  
US-09-813-453A-17  
; Sequence 17, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 17  
; LENGTH: 233  
; TYPE: PRT  
; ORGANISM: Bacillus subtilis  
US-09-813-453A-17

Query Match 37.9%; Score 492; DB 9; Length 233;  
Best Local Similarity 45.2%; Pred. No. 7.3e-39;  
Matches 95; Conservative 41; Mismatches 72; Indels 2; Gaps 1;  
Qy 1 MLLAIEQGNNTMTMFAIHDGASWVAQWRSATESSTRTADEYVWVLSQLLSMQGLGFRDAIDAV 60  
Db 1 MLLVIDVGNNTNVLGVYHDGKLEYHWRIETSRHKTDEFEGMILRSLFDHSGLMFEQIDGI 60  
Qy 61 IISVVVPQSIFNLNLSRRYFNVEPLVIGENAKGLIDVRIEKPSEAGADRLVNAIGAAMV 120  
Db 61 IISVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYNPNKQVGADRVNAVAAIHL 120  
Qy 121 YPGPLVIDSGTATTFDIAADGAFEGGIIAPGINLSMQALHEAAKLPRIAIQRPAGNR 180  
Db 121 YGNPLIVDFGTATTCYCIDENKQYMGGAIPGITISTEALYSRAAKLPRIETRP--DN 178

Db 179 IIGKNTVSAMQSGILFGYGVQVEGIVKRMKWAQDLKVIATGGGLAPLIANESDCIDVD 238  
QY 241 SDLTIRGLLEIYRNTI 257  
Db 239 PFLTGLKLELIYERNRV 255

RESULT 5  
US-09-813-453A-49  
; Sequence 49, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR FILING DATE: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 49  
; LENGTH: 258  
; TYPE: PRT  
; ORGANISM: Bacillus stearothermophilus  
US-09-813-453A-49

Query Match 44.3%; Score 575; DB 9; Length 258;  
Best Local Similarity 44.1%; Pred. No. 1.1e-46;  
Matches 113; Conservative 52; Mismatches 89; Indels 2; Gaps 1;  
QY 1 MLLAIEQGNNTMTFAIHDGASWVAQWRSATSTRTADEYVWLSQLLSMOGLGFRAIDAV 60  
Db 1 MIFVLDVGNNTVGLGYDDELKHHWRIETSRSKTEDEYGMKALKLNHVGLQFSDIRGI 60  
QY 61 IISVVVQSIFNLNLSRRYFNVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAMV 120  
Db 61 IISVVVPPIMFALERMCLEKHYFKIKPLVGPVGTGDIKIDYDNPREVGDRIVNAVAGIHL 120  
QY 121 YPGPLVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMQALHEAAKLPRIAIORPAGNR 180  
Db 121 YGSPLIIVDFGTATTYCYINEHKQYMGGAIAPIGIMISTEALFARAALKPRIEIRP--DD 178  
QY 181 IVGTDVTSAMQSGVFWGYISLIEGLVARIKAERGEPTMTVIATGGVASLFEGATDSIDHFD 240  
Db 179 IIGKNTVSAMQAGILYGVQVEGIVSRMKAKSKIPPVVIATGGGLAPLIASESDIIDVVD 238  
QY 241 SDLTIRGLLEIYRNT 256  
Db 239 PFLTGLKLLYEKNT 254

RESULT 6  
US-09-813-453A-45  
; Sequence 45, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR FILING DATE: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21

; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 45  
; LENGTH: 262  
; TYPE: PRT  
; ORGANISM: Bacillus anthracis  
US-09-813-453A-45

Query Match 43.6%; Score 566; DB 9; Length 262;  
Best Local Similarity 42.0%; Pred. No. 8.3e-46;  
Matches 107; Conservative 55; Mismatches 91; Indels 2; Gaps 1;  
QY 1 MLLAIEQGNNTMTFAIHDGASWVAQWRSATSTRTADEYVWLSQLLSMOGLGFRAIDAV 60  
Db 1 MIFVLDVGNNTNAVGLGVFEEGELRQHRMETDRHKTEDEYGLVKQLLEHEGLSFEDVKGI 60  
QY 61 IISVVVQSIFNLNLSRRYFNVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAMV 120  
Db 61 IISVVVPPIMFALERMCLEKHYFKIKPLVGPVGTGDIKIDYDNPREVGDRIVNAVAGIHL 120  
QY 121 YPGPLVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMQALHEAAKLPRIAIORPAGNR 180  
Db 121 YGSPLIIVDFGTATTYCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETKPSS-- 178  
QY 181 IVGTDVTSAMQSGVFWGYISLIEGLVARIKAERGEPTMTVIATGGVASLFEGATDSIDHFD 240  
Db 179 VVGKNTVSAMQSGILYGVQVEGIVKRMKEAKQEPKVIATGGGLAKLISESNVIDVD 238  
QY 241 SDLTIRGLLEIYRNRN 255  
Db 239 PFLTGLKGLYLYERN 253

RESULT 7  
US-09-813-453A-7  
; Sequence 7, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 255  
; TYPE: PRT  
; ORGANISM: Geobacter sulfurreducens  
US-09-813-453A-7

Query Match 43.2%; Score 561; DB 9; Length 255;  
Best Local Similarity 44.7%; Pred. No. 2.4e-45;  
Matches 114; Conservative 50; Mismatches 89; Indels 2; Gaps 1;  
QY 1 MLLAIEQGNNTMTFAIHDGASWVAQWRSATSTRTADEYVWLSQLLSMOGLGFRAIDAV 60  
Db 1 MLLVIDVGNNTNVLGIYDGERLVRDWRVSTDKARTTDEYGLINELFRLAGLGLDQIRAV 60  
QY 61 IISVVVQSIFNLNLSRRYFNVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAMV 120  
Db 61 IISVVVPLTGVLERLSLIGYFGMRPLVVGPGIKTGMPIQYDNPREVGDRIVNAVAGYEK 120  
QY 121 YPGPLVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMQALHEAAKLPRIAIORPAGNR 180  
Db 121 YRTSLIIVDFGTATFDYVNRKGEYCGGAIAPLGVISTEALFQRASKLPVVDIIRPSA-- 178



Db 181 IVGTDTVSAMQSGVFWGYISLIEGLVARIKAERGEPMVTIATGGVASLFEAGTDSIDHFD 240  
QY 241 SDLTIRGLLEIYRRNTIAES 260  
Db 241 SDLTIRGLLEIYRRNTIAES 260

RESULT 2  
US-09-813-453A-6  
; Sequence 6, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OG2-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 258  
; TYPE: PRT  
; ORGANISM: Rhodobacter capsulatus  
US-09-813-453A-6

Query Match 50.7%; Score 658; DB 9; Length 258;  
Best Local Similarity 50.8%; Pred. No. 1.6e-54;  
Matches 130; Conservative 48; Mismatches 74; Indels 4; Gaps 3;  
QY 1 MLLAIEQGNNTMTFAIHDGASWVAQWRSATESRTRTADEYVVWLSQLLSMQGLGFRAIDAV 60  
Db 1 MLLCIDCGNTNTVFSVWDGTDFAATWRIATDHRRTADEYFVWLTMLQKLGQR-ISEA 59  
QY 61 IISVVVPQSIFNLRNLSRRYFNVPEPLVIGE-NAKLGIDVRIEKPSEAGADRLVNAIGAAM 119  
Db 60 IISSTAPRVFNLRVLCNRYFDCRPYVVGKPCGCELPVAPRVDPTTVPDRLVNTVAGYD 119  
QY 120 YPGPLVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMQALHEAAAKLPRIAIQRPAGN 179  
Db 120 RHGDLIVDFGTATTFDVAADGAYIGGVIAAGVNLSEALHMAAALPHVDVTKPGQ- 178  
QY 180 RIVGTDTVSAMQSGVFWGYISLIEGLVARIKAERGEPMVTIATGGVASLFEAGTDSIDHF 239  
179 -VIGTNTVACIQSGVYWGVIYGLVEGIVRQIRMERDRPMKVIAATGGLASLFDLGFDFDKV 237  
QY 240 DSDLTIRGLLEIYRRN 255  
Db 238 EDDLTMHGLRLIFDYN 253

RESULT 3  
US-09-813-453A-47  
; Sequence 47, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OG2-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 47  
; LENGTH: 254  
; TYPE: PRT  
; ORGANISM: Bacillus halodurans  
US-09-813-453A-47  
Query Match 46.0%; Score 598; DB 9; Length 254;  
Best Local Similarity 43.9%; Pred. No. 7.4e-49;  
Matches 112; Conservative 61; Mismatches 80; Indels 2; Gaps 1;  
QY 1 MLLAIEQGNNTMTFAIHDGASWVAQWRSATESRTRTADEYVVWLSQLLSMQGLGFRAIDAV 60  
Db 1 MILVIDGNTNTVLGVYQDETLYVHHWRLATSROKTEDEYAMTVRSLEFDHAGLQFQDIDGI 60  
QY 61 IISVVVPQSIFNLRNLSRRYFNVPEPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAMV 120  
Db 61 VISSVVPMMFSLQCMCKKYFHTPMIGPKITGLNICYDNPKYEGADRVNVAIAIEL 120  
QY 121 YPGPLVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMQALHEAAAKLPRIAIQRPAGNR 180  
Db 121 YGPAIVVDFGTATTYCLINEKKQYAGGVIAAGIMISTEALYHRASKLPRIEIAKP--KQ 178  
QY 181 IVGTDTVSAMQSGVFWGYISLIEGLVARIKAERGEPMVTIATGGVASLFEAGTDSIDHFD 240  
Db 179 VVGNTIDSMQSGIFYGVYVQVDGVVVKRMAQAESEPKVIATGGLAKLIGTSETIDVID 238  
QY 241 SDLTIRGLLEIYRRN 255  
Db 239 SFTLTKGLLIYKKN 253

RESULT 4  
US-09-813-453A-2  
; Sequence 2, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OG2-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 258  
; TYPE: PRT  
; ORGANISM: Bacillus subtilis  
US-09-813-453A-2

Query Match 44.4%; Score 577; DB 9; Length 258;  
Best Local Similarity 44.4%; Pred. No. 7.4e-47;  
Matches 114; Conservative 50; Mismatches 91; Indels 2; Gaps 1;  
QY 1 MLLAIEQGNNTMTFAIHDGASWVAQWRSATESRTRTADEYVVWLSQLLSMQGLGFRAIDAV 60  
Db 1 LLLVIDGNTNTVLGVYHDKLEYHWRITSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60  
QY 61 IISVVVPQSIFNLRNLSRRYFNVPEPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAMV 120  
Db 61 IISVVPPIMFALERMCTKYFHIQVIGPMKTNLNIKYDNPKYEGADRVNVAIAIHL 120  
QY 121 YPGPLVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMQALHEAAAKLPRIAIQRPAGNR 180  
Db 121 YGNPLIVDFGTATTYCYIDENKQYMGGAIPGITISTEALYSRAAKLPRIETRP--DN 178  
QY 181 IVGTDTVSAMQSGVFWGYISLIEGLVARIKAERGEPMVTIATGGVASLFEAGTDSIDHFD 240

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

'OM protein - protein search, using sw model

Run on: June 24, 2003, 22:24:12 ; Search time 14.4494 Seconds  
(without alignments)  
1947.059 Million cell updates/sec

Title: US-09-813-453A-51  
Perfect score: 1299  
Sequence: 1 MLLAIEQGNTMTFAIHDGA.....SDLTIRGLLEIYRNTIAES 260

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1299	100.0	260	9	US-09-813-453A-51
2	658	50.7	258	9	US-09-813-453A-6
3	598	46.0	254	9	US-09-813-453A-47
4	577	44.4	258	9	US-09-813-453A-2
5	575	44.3	258	9	US-09-813-453A-49
6	566	43.6	262	9	US-09-813-453A-45
7	561	43.2	255	9	US-09-813-453A-7
8	523.5	40.3	256	9	US-09-813-453A-55
9	517	39.8	265	9	US-09-813-453A-4
10	492	37.9	233	9	US-09-813-453A-17
11	478	36.8	219	9	US-09-813-453A-57
12	460.5	35.5	250	9	US-09-813-453A-3
13	396.5	30.5	262	9	US-09-813-453A-8
14	382.5	29.4	272	9	US-09-813-453A-5
15	378.5	29.1	272	9	US-09-712-363-276
16	359.5	27.7	246	9	US-09-813-453A-9
17	337.5	26.0	273	9	US-09-813-453A-10
18	307.5	23.7	212	9	US-09-813-453A-59
19	261	20.1	257	9	US-09-813-453A-53

20	227	17.5	241	9	US-09-813-453A-63	Sequence 63, Appl
21	198	15.2	244	9	US-09-813-453A-41	Sequence 41, Appl
22	189.5	14.6	262	9	US-09-813-453A-11	Sequence 11, Appl
23	185	14.2	460	9	US-09-813-453A-39	Sequence 39, Appl
24	175	13.5	592	9	US-09-813-453A-22	Sequence 22, Appl
25	172	13.2	592	9	US-09-813-453A-43	Sequence 43, Appl
26	160	12.3	248	9	US-09-813-453A-20	Sequence 20, Appl
27	153.5	11.8	249	9	US-09-813-453A-61	Sequence 61, Appl
28	147	11.3	257	9	US-09-813-453A-13	Sequence 13, Appl
29	141	10.9	249	9	US-09-813-453A-70	Sequence 70, Appl
30	140.5	10.8	242	9	US-09-813-453A-65	Sequence 65, Appl
31	134.5	10.4	267	9	US-09-813-453A-15	Sequence 15, Appl
32	125.5	9.7	229	9	US-09-813-453A-12	Sequence 12, Appl
33	111	8.5	209	9	US-09-813-453A-21	Sequence 21, Appl
34	98	7.5	1536	9	US-10-092-880-2	Sequence 2, Appl
35	94.5	7.3	448	9	US-09-738-626-3991	Sequence 3991, Ap
36	93	7.2	491	9	US-09-738-626-4901	Sequence 4901, Ap
37	92	7.1	511	9	US-09-738-626-4770	Sequence 4770, Ap
38	89	6.9	206	9	US-09-997-816-4	Sequence 4, Appl
39	88	6.8	255	10	US-09-847-637B-9	Sequence 9, Appl
40	87.5	6.7	343	9	US-10-168-066-4	Sequence 4, Appl
41	87.5	6.7	923	10	US-09-815-242-11905	Sequence 11905, A
42	87	6.7	242	10	US-09-815-242-10582	Sequence 10582, A
43	87	6.7	638	10	US-09-815-242-13713	Sequence 13713, A
44	85	6.5	398	9	US-09-738-626-6322	Sequence 6322, Ap
45	85	6.5	402	10	US-09-838-564A-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-09-813-453A-51  
; Sequence 51, Application US/09813453A  
; Patent No. US20020168681A1

GENERAL INFORMATION:

; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 51

; LENGTH: 260

; TYPE: PRT

; ORGANISM: Caulobacter crescentus

US-09-813-453A-51

Query Match 100.0%; Score 1299; DB 9; Length 260;  
Best Local Similarity 100.0%; Pred. No. 2.9e-115;  
Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MLLAIEQGNTMTFAIHDGASVWAQWRSATES	TRTADYVVWLSQLLSMQGLGFRAIDAV	60
Db	1	MLLAIEQGNTMTFAIHDGASVWAQWRSATES	TRTADYVVWLSQLLSMQGLGFRAIDAV	60
QY	61	IISVVVPQSIFNLRLNSRRYFNVEPLVIGENAKL	GIDVRIEKPSEAGADRLVNAIGAMV	120
Db	61	IISVVVPQSIFNLRLNSRRYFNVEPLVIGENAKL	GIDVRIEKPSEAGADRLVNAIGAMV	120
QY	121	YPGLVVIDSGTATFDIVAADGAFEGGIIAPGINLS	MQALHEAAKLPRIAIQRPAGNR	180
Db	121	YPGLVVIDSGTATFDIVAADGAFEGGIIAPGINLS	MQALHEAAKLPRIAIQRPAGNR	180
QY	181	IVGTDTVSAMQSGVFWGYISLIEGLVARIKAE	RGEPMTVIATGGVASLFEGATDSIDHFD	240
Db	181	IVGTDTVSAMQSGVFWGYISLIEGLVARIKAE	RGEPMTVIATGGVASLFEGATDSIDHFD	240

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:49:47 ; Search time 10.5489 Seconds  
(without alignments)  
725.188 Million cell updates/sec

Title: US-09-813-453A-51  
Perfect score: 1299  
Sequence: 1 MLLAIEQGNTWTFMFAHDGA.....SDLTIRGLLEIYRRNTIAES 260

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	98	7.5	1095	4	US-09-206-942-69
2	98	7.5	1536	1	US-08-038-682-2
3	98	7.5	1536	1	US-08-302-832-2
4	98	7.5	1536	2	US-08-530-198-2
5	98	7.5	1536	2	US-08-469-880-2
6	98	7.5	1536	2	US-08-728-470-2
7	98	7.5	1536	2	US-08-617-697-2
8	98	7.5	1536	4	US-08-719-641-2
9	98	7.5	1536	4	US-09-206-942-67
10	90	6.9	1220	4	US-09-206-942-28
11	89.5	6.9	1226	4	US-09-206-942-26
12	89.5	6.9	1891	2	US-08-804-227C-12
13	85	6.5	1489	6	5183745-2
14	85	6.5	1706	4	US-08-669-785-2
15	85	6.5	1794	6	5183745-6
16	85	6.5	1705	4	US-08-669-785-4
17	84.5	6.4	659	4	US-08-894-818B-1
18	83.5	6.4	659	4	US-09-445-472-12
19	83.5	6.4	1180	4	US-09-206-942-65
20	83	6.4	1188	4	US-09-206-942-63
21	83	6.4	3170	4	US-09-036-987A-4
22	82.5	6.4	3170	4	US-09-370-700-4
23	82.5	6.3	1095	4	US-09-206-942-45
24	82	6.3	1101	4	US-09-206-942-43
25	82	6.3	4150	4	US-09-428-517-2
26	81.5	6.3	3729	2	US-08-804-227C-4
27	81	6.2			

28	80	6.2	352	4	US-09-504-358-18	Sequence 18, Appl
29	80	6.2	352	4	US-09-954-314-18	Sequence 18, Appl
30	79.5	6.1	355	4	US-08-818-112-79	Sequence 79, Appl
31	79.5	6.1	355	4	US-08-818-111-80	Sequence 80, Appl
32	79.5	6.1	355	4	US-09-056-556-79	Sequence 79, Appl
33	79.5	6.1	355	4	US-09-072-596-80	Sequence 80, Appl
34	79.5	6.1	395	2	US-08-404-531B-3	Sequence 3, Appl
35	79.5	6.1	395	3	US-08-476-900A-3	Sequence 3, Appl
36	79.5	6.1	395	3	US-08-488-546A-3	Sequence 3, Appl
37	79.5	6.1	922	4	US-08-460-269C-6	Sequence 6, Appl
38	79.5	6.1	1580	4	US-08-726-320-1	Sequence 1, Appl
39	79.5	6.1	1580	4	US-09-208-716-1	Sequence 1, Appl
40	79.5	6.1	1581	4	US-08-726-320-3	Sequence 3, Appl
41	79.5	6.1	1581	4	US-09-208-716-3	Sequence 3, Appl
42	79	6.1	1544	4	US-09-413-814-46	Sequence 46, Appl
43	78.5	6.0	643	4	US-09-178-252-25	Sequence 25, Appl
44	78.5	6.0	1186	4	US-09-178-252-23	Sequence 23, Appl
45	78.5	6.0	1227	1	US-08-448-170-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1  
US-09-206-942-69  
; Sequence 69, Application US/09206942  
; Patent No. 6432669  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High  
; TITLE OF INVENTION: Molecular Weight Proteins  
; FILE REFERENCE: 1038-861 MIS:jb  
; CURRENT APPLICATION NUMBER: US/09/206,942  
; CURRENT FILING DATE: 1998-12-08  
; EARLIER APPLICATION NUMBER: 09/167,568  
; EARLIER FILING DATE: 1998-10-07  
; NUMBER OF SEQ ID NOS: 95  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 69  
; LENGTH: 1095  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-09-206-942-69

Query Match	7.5%	Score 98;	DB 4;	Length 1095;
Best Local Similarity	29.0%	Pred. No. 0.035;		
Matches	45;	Conservative	22;	Mismatches 58; Indels 30; Gaps 7;
Qy	126	VVIDSGT---ATT--FDIVAADGAFEGGIIAPGINLSMQALHEAAKLPRI-----	171	
Db	706	ITTKTGTTINATTGNVEITAQTGSILGGIESSSGSVTLTAT-EGALAVSNISGNTVTVTA	764	
Qy	172	---AIQRPAGNRIVGDTV-SAMQSGVFWGYIS-----LIEGLV----ARIKAERGE	216	
Db	765	NSGALTLAGSTIKGTESVTTSSQSGDIGGTISGGTVEVKATESLTQSNKIKATTGEA	824	
Qy	217	MTVIATGGVASLFEGATDSIDHFDSDLTIRGLLEI	251	
Db	825	NVTSATGTIGGTISGNTVNVNANAGDLTVGNAGEI	859	

RESULT 2  
US-08-038-632-2  
; Sequence 2, Application US/08038682  
; Patent No. 5549897  
; GENERAL INFORMATION:  
; APPLICANT: BARENKAMP, STEPHEN J  
; APPLICANT: ST. GEME III, JOSEPH W  
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS  
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS  
; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; STREET: Bldg. 1  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/038,682  
; FILING DATE: 16-MAR-1993  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BERKSTRESSER, JERRY W  
; REGISTRATION NUMBER: 22,651  
; REFERENCE/DOCKET NUMBER: 1038-293  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0813  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1536 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-038-682-2

Query Match 7.5%; Score 98; DB 1; Length 1536;  
Best Local Similarity 29.0%; Pred. No. 0.059;  
Matches 45; Conservative 22; Mismatches 58; Indels 30; Gaps 7;

QY 126 VVIDSGT---ATT--FDIVAADGAFEGGIIAPGINLSMOALHEAAKLPR1----- 171  
Db 1147 ITTKGTGINATTGNVEITAGTGSILGGIESSSGSVTLTAT-EGALAVSNISGNTVTVTA 1205

QY 172 ---AIQRPAGNRIVGTDIV-SAMQSGVFWGYIS-----LIEGLV---ARIKAERGE 216  
Db 1206 NSGALTTLAGSTIKGTESVTTSSQSGDIGGTISGGTVEVKATESLTQNSKIKATTGEA 1265

QY 217 MTVIATGGVASLFEAGTDSIDHFDSDLTIRGLLEI 251  
Db 1266 NVTSATGTIGGTISGNTVNTANAGDLTVGNAGAEI 1300

ESULT 3  
US-08-302-832-2  
; Sequence 2, Application US/08302832  
; Patent No. 5603938  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd.  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; STREET: Bldg. 1  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/302,832  
; FILING DATE: 16-SEP-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9205704.1  
; FILING DATE: 16-MAR-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US pct/us93/02166  
; FILING DATE: 16-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Berkstresser, Jerry W  
; REGISTRATION NUMBER: 22,651  
; REFERENCE/DOCKET NUMBER: 1038-404  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0813  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1536 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-302-832-2

Query Match 7.5%; Score 98; DB 1; Length 1536;  
Best Local Similarity 29.0%; Pred. No. 0.059;  
Matches 45; Conservative 22; Mismatches 58; Indels 30; Gaps 7;

QY 126 VVIDSGT---ATT--FDIVAADGAFEGGIIAPGINLSMOALHEAAKLPR1----- 171  
Db 1147 ITTKGTGINATTGNVEITAGTGSILGGIESSSGSVTLTAT-EGALAVSNISGNTVTVTA 1205

QY 172 ---AIQRPAGNRIVGTDIV-SAMQSGVFWGYIS-----LIEGLV---ARIKAERGE 216  
Db 1206 NSGALTTLAGSTIKGTESVTTSSQSGDIGGTISGGTVEVKATESLTQNSKIKATTGEA 1265

QY 217 MTVIATGGVASLFEAGTDSIDHFDSDLTIRGLLEI 251  
Db 1266 NVTSATGTIGGTISGNTVNTANAGDLTVGNAGAEI 1300

RESULT 4  
US-08-530-198-2  
; Sequence 2, Application US/08530198  
; Patent No. 5869065  
; GENERAL INFORMATION:  
; APPLICANT: BARENKAMP, STEPHEN J  
; APPLICANT: ST. GEME III, JOSEPH W  
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS  
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; STREET: Bldg. 1  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/530,198  
; FILING DATE: 13-DEC-1995  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BERKSTRESSER, JERRY W  
; REGISTRATION NUMBER: 22,651  
; REFERENCE/DOCKET NUMBER: JWB-1186



TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1536 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-530-198-2

Query Match 7.5%; Score 98; DB 2; Length 1536;  
Best Local Similarity 29.0%; Pred. No. 0.059;  
Matches 45; Conservative 22; Mismatches 58; Indels 30; Gaps 7;

QY 126 VVIDSGT---ATT--FDIVAADGAFEGGIIAPGINLSMQALHEAAKLPRI----- 171  
Db 1147 ITTKGTGINATTGNVEITAGTGSILGGIESSSGSVTLTAT-EGALAVSNISGNTVTVTA 1205  
QY 172 ---AIQRPAGNRIVGTDV-SAMQSGVFWGYIS-----LIEGLV----ARIKAERGE 216  
Db 1206 NSGALTTLAGSTIKGTESVTTSSQSGDIGGTISGGTVEVKATESLTQSNKIKATTGEA 1265

QY 217 MTVIATGGVÀSLFEGATDSIDHFDSDLTIRGLLEI 251  
Db 1266 NVTSATGTIGGTISGNTVNVNANAGDLTVGNCAEI 1300

RESULT 5

US-08-469-880-2  
Sequence 2, Application US/08469880  
Patent No. 5876733

GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J.  
TITLE OF INVENTION: High Molecular Weight Surface Proteins  
TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
STREET: Bldg. 1  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,880  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9205704.1  
FILING DATE: 16-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 16-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-516 MIS:vg  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:

LENGTH: 1536 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-469-880-2

Query Match 7.5%; Score 98; DB 2; Length 1536;  
Best Local Similarity 29.0%; Pred. No. 0.059;  
Matches 45; Conservative 22; Mismatches 58; Indels 30; Gaps 7;

QY 126 VVIDSGT---ATT--FDIVAADGAFEGGIIAPGINLSMQALHEAAKLPRI----- 171  
Db 1147 ITTKGTGINATTGNVEITAGTGSILGGIESSSGSVTLTAT-EGALAVSNISGNTVTVTA 1205  
QY 172 ---AIQRPAGNRIVGTDV-SAMQSGVFWGYIS-----LIEGLV----ARIKAERGE 216  
Db 1206 NSGALTTLAGSTIKGTESVTTSSQSGDIGGTISGGTVEVKATESLTQSNKIKATTGEA 1265  
QY 217 MTVIATGGVASLFEGATDSIDHFDSDLTIRGLLEI 251  
Db 1266 NVTSATGTIGGTISGNTVNVNANAGDLTVGNCAEI 1300

RESULT 6

US-08-728-470-2  
Sequence 2, Application US/08728470  
Patent No. 5928651

GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J  
TITLE OF INVENTION: High Molecular Weight Surface Proteins  
TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
STREET: Bldg. 1  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/728,470  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9205704.1  
FILING DATE: 16-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-633  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1536 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE:

US-08-728-470-2

Query Match 7.5%; Score 98; DB 2; Length 1536;  
Best Local Similarity 29.0%; Pred. No. 0.059;  
Matches 45; Conservative 22; Mismatches 58; Indels

```

126 QY 126 VVIDSGT---ATT--FDIVAADGAFEGGIIAPGINLSMQALHEAAAKLPRI----- 171
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1147 Db 1147 ITTKTGTINATTGNVEITAQTGSILGGIESSGSVTLTAT-EGALAVSNISGNTVTVTA 1205
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
172 QY 172 ---AIQRPAGNRIVGTDTV-SAMQSGVFWGYIS-----LIEGLV---ARIKAERGE 216
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1206 Db 1206 NSGALTTLAGSIKGTESVTTSSQSGDIGCTTSGGTVEVKATESLTQTQNSKIKATTGEA 1265
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
217 QY 217 MTVIATGGVASLFEGATDSIDHFDSDLTIRGLLEI 251
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1266 Db 1266 NVTSATGTIGGTISGNTVNVVTANAGDLTVNGAEI 1300
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

```

## RESULT 7

US-08-617-697-2  
Sequence 2, Application US/08617697  
Patent No. 5977336  
GENERAL INFORMATION:  
APPLICANT: Barenkamp, Stephen J  
TITLE OF INVENTION: High Molecular Weight Surface Proteins  
TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
STREET: Bldg. 1  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/617,697  
FILING DATE: 01-APR-1996  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 05-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-557  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1536 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-617-697-2

```

Query Match          7.5%; Score 98; DB 2; Length 1536;
Best Local Similarity 29.0%; Pred. No. 0.059;
Matches 45; Conservative 22; Mismatches 58; Indels 30; Gaps 7

QY      126 VVDSGT---ATT--FDIVAADGAFEGGIIAPGINLSMQALHEAAAKLPRI-----171
       :   ::|||    :|||   :|||   :|||   :|||   :|||   :|||   :|||

```

Db	1147	ITTKTGT	TINAT	TGNVE	ITAQ	TG	SILG	GI	ESSG	SV	TL	TAT	-	EG	AL	AV	SN	IS	GN	T	V	T	V	T	A	1205																															
QY	172	--	AI	Q	P	A	G	N	R	I	V	G	T	D	T	V	-	S	A	M	Q	S	G	V	F	W	G	Y	I	S	---	L	I	E	G	L	V	---	A	R	I	K	A	E	R	G	E	P	216								
Db	1206	NSG	AL	T	T	L	A	G	S	T	I	K	G	T	E	S	V	T	T	S	S	Q	S	G	D	I	G	G	T	I	S	G	G	T	V	E	K	A	T	E	S	L	T	T	Q	S	N	K	I	K	A	T	T	G	E	A	1265
QY	217	M	T	V	I	A	T	G	V	A	S	L	F	E	G	A	T	D	S	I	D	H	F	D	S	D	L	T	I	R	G	L	L	E	I	251																					
Db	1266	N	V	T	S	A	T	G	T	I	G	S	T	I	S	G	N	T	V	N	V	T	A	N	A	G	D	L	T	V	G	N	G	A	E	I	1300																				

## RESULT 8

US-08-719-641-2  
; Sequence 2, Application US/08719641  
; Patent No. 6218141  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; TITLE OF INVENTION: Of No. 6218141-Typeable Haemophilus  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd.  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/719,641  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/302,832  
; FILING DATE: 16-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US PCT/US93/02166  
; FILING DATE: 16-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9205704.1  
; FILING DATE: 16-MAR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Berkstresser, Jerry W  
; REGISTRATION NUMBER: 22,651  
; REFERENCE/DOCKET NUMBER: 1038-625  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0813  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1536 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE:  
US-08-719-641-2

	Query Match	7.5%; Score 98; DB 4; Length 1536;
	Best Local Similarity 29.0%; Pred. No. 0.059;	
	Matches 45; Conservative 22; Mismatches 58; Indels 30; Gaps 7;	
QY	126 VVIDSGT---ATT--FDIVAADGAFEGGIIAPGINLSMQALHEAAKLPRI-----	171
	: :    :    :    :    :    :    :    :    :	
Db	1147 ITTKTGTINATTGNVEITAQTGSIILGGIESSSGSVTLTAT-EGALAVSNISGNTVTVTA	1205
QY	172 ---AIQRPAGNRIVGTDTV-SAMQSGVFWGYIS-----LIEGLV----ARIKAERGE	216
	: :    :    :    :    :    :    :    :	
Db	1206 NSGALTTLAGSTIKGTESVTTSSOSGDIGGTISGGTVEVKATESLTTOSNSKIKATTGEA	1265



```

;
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rosteck, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1891 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
;
; US-08-804-227C-12

```

```

Query Match 6.9%; Score 89.5; DB 2; Length 1891;
Best Local Similarity 27.7%; Pred. No. 0.8;
Matches 41; Conservative 18; Mismatches 62; Indels 27; Gaps 6;

QY 99 RIEKPSEAG-ADRLVNAIGAAMVYPGPLVVIDSGTATTFDIVAADGAFEGGIIAPGINLS 157
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
1073 RTETPDRTGTLAARLAELARSPEGLAGVLLLPDSGGAA----VAGHPGLDQGTAA--VLLT 1126
: | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :

QY 158 MQALHEAAAKLPRIAQRPA-----GNRIVGTDTVSAMQSGVFW-GYISLIEG 204
: | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
1127 IQALTDAAVRAPLWVTVRGAVAVGSGEVPICAVGARVWGLGRVAALVPPVQMGGLVDVAVG 1186

' 205 L-----VARIKAERGEPTVIATGGV 225
| : | | | : | | |

Db 1187 AGVRERRRVGVVAGGEGEDQVAVRGGV 1214

```

```

RESULT 13
US-08-804-198-6
; Sequence 6, Application US/08804198
; Patent No. 5945320
;
; GENERAL INFORMATION:
; APPLICANT: Burgett, Stanley G.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rao, Nagaraja R.
; APPLICANT: Richardson, Mark A.
; APPLICANT: Rosteck, Paul R., Jr.
; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PAUL R. CANTRELL 1138
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285

```

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,198
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CANTRELL, PAUL R.
; REGISTRATION NUMBER: 36,470
; REFERENCE/DOCKET NUMBER: P9113
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3885
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1891 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
;
; US-08-804-198-6

```

```

Query Match 6.9%; Score 89.5; DB 2; Length 1891;
Best Local Similarity 27.7%; Pred. No. 0.8;
Matches 41; Conservative 18; Mismatches 62; Indels 27; Gaps 6;

QY 99 RIEKPSEAG-ADRLVNAIGAAMVYPGPLVVIDSGTATTFDIVAADGAFEGGIIAPGINLS 157
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
1073 RTETPDRTGTLAARLAELARSPEGLAGVLLLPDSGGAA----VAGHPGLDQGTAA--VLLT 1126
: | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :

QY 158 MQALHEAAAKLPRIAQRPA-----GNRIVGTDTVSAMQSGVFW-GYISLIEG 204
: | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
1127 IQALTDAAVRAPLWVTVRGAVAVGSGEVPICAVGARVWGLGRVAALVPPVQMGGLVDVAVG 1186

QY 205 L-----VARIKAERGEPTVIATGGV 225
| : | | | : | | |

Db 1187 AGVRERRRVGVVAGGEGEDQVAVRGGV 1214

```

```

RESULT 14
5183745-2
; Patent No. 5183745
; APPLICANT: DANCHIN, ANTOINE; GLASER, PHILLIPPE; KRIN, EVELYN;
; BARZU, OCTAVIEN; LADANT, DANIEL; ULLMAN, AGNES
; TITLE OF INVENTION: ADENYL CYCLASE DERIVATIVES AND THEIR
; BIOLOGICAL USES
; NUMBER OF SEQUENCES: 13
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/426,541
; FILING DATE: 25-OCT-1989
; SEQ ID NO: 2:
; LENGTH: 1489
5183745-2

```

```

Query Match 6.5%; Score 85; DB 6; Length 1489;
Best Local Similarity 20.0%; Pred. No. 1.8;
Matches 55; Conservative 29; Mismatches 113; Indels 78; Gaps 8;

QY 48 SMQGLGFRAI-----DAVISSVVPQSIFNLRLNLSRRYFNVPEPLVIGENAKLGDVRIE 101
| : | | | : | : | : | : | : | : | : | : | : | : | : | : | :
407 SLDGVGSRFSFSLGEVSDMAAVEAAELEMTQVHLHAGARQDDAEPPGVSGASAHWQRALQG 466

QY 102 KPSEAGADRLVNAIGAAMVYPG-----PLVVIDSQTATTFDIVAADGAFE- 146
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
467 AQAVAAAQRLVHAI-ALMTQFGRAGSTNTPQEAASLSAAVFGLEASSAVAETVSGFFRG 525

QY 147 -----GGIIAPGINLSMQALHEAAAKLPRIAQRPAAGNRIVGTDTVSA 189
| | | | | : | | | : | | | : | | | : | | | : | | | : | | | :
526 SSRWAGGFGVAGGAMALGGGIAAAGAGMSLTDDA-----PAGKAAAGAEIAL 574

QY 190 MQSGVFWGYISLIEGLVARIKAERGEPTVIATGGVASLFEGA----- 232
: | | | : | | : | | | : | | | : | | | : | | | : | | | : | | | :

```



```

Db          6531 HYAQDLKLAQESSAYGYEGDALLAQLYRDKTAAE 6665

Search completed: June 24, 2003, 22:10:43
Job time : 11.5489 secs

```

## RESULT 15

US-08-669-785-2  
; Sequence 2, Application US/08669785  
; Patent No. 6309648  
; GENERAL INFORMATION:  
; APPLICANT: Betsou, Fotini  
; APPLICANT: Sebo, Peter  
; APPLICANT: Guiso, Nicole  
; TITLE OF INVENTION: Protective Epitopes Of Adenyl  
; CYCLASE-HAEMOLYSIN(AC-HLY), Their Application To  
; TITLE OF INVENTION: The Treatment Or To The Prevention Of Bordetella Infections.  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
; STREET: 1300 I Street, N.W., Suite 700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/669,785  
; FILING DATE: 27-JUN-1996  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meyers, Kenneth J.  
; REGISTRATION NUMBER: 25,146  
; REFERENCE/DOCKET NUMBER: 02356.0072-00000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 408-4000  
; TELEFAX: (202) 408-4400  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1706 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: Protein  
US-08-669-785-2

Query Match	6.5%;	Score 85;	DB 4;	Length 1706;
Best Local Similarity	20.0%;	Pred. NO. 2.3;		
Matches	55; Conservative	29; Mismatches	113; Indels	78; Gaps
QY	48	SMOGLGFRAI-----DAVISSVVPQSFIFNLNLSRRYFNVEPLVIGENAKLIGIDVRIE	101	
Db	406	SLDGVGSRFSLSGEVSDMAAVEAAEELEMTQVLHAGARQDDAEPGVSGASAHWGQALQG	465	
QY	102	KPSEAGADRLVNAIGAAMVYPG-----PLVVIDSGTATTFDIVAADGAFE--	146	
Db	466	AQAVAAAQRLVHAI-ALMTQFGRAGSTNTPOEAAASLSAAVFGLGEASSAVAETVSGFFRG	524	
QY	147	-----GGIIAPGINLSMQALHEAAAKLPRIAIOQPAGNRIVGTDVSA	189	
Db	525	SSRWAGGFGVAGGAMALGGGIAAAVAGAGMSLTDDA-----PAGQKAAAGABIAL	573	
QY	190	MQSGVFWGVYISLIEGLVARIKAERGEPMVTIATGGVASLFEGA-----	232	
Db	574	QLTG---GVVELASSIALALAAARGVTSGLQVAGASAGAAAAGALAAALSPMEIYGLVQQS	630	
QY	233	--TDSIDHFDSDLTIRG-----LLEIYRRNTIAE	259	

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:03:27 ; Search time 10.7262 Seconds  
(without alignments)  
2330.267 Million cell updates/sec

Title: US-09-813-453A-51  
Perfect score: 1299  
Sequence: 1 MLLAIEQGNTNTMFAIHGDA.....SDLTIRGLLEIYRRNTIAES 260

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1299	100.0	261	2	B87489	transcription acti
2	598	46.0	254	2	F83660	hypothetical prote
3	523	40.3	259	2	AF1102	conserved hypothe
4	519	40.0	273	2	E97293	probable transcrip
5	517	39.8	265	2	T36391	hypothetical prote
6	516	39.7	259	2	AF1464	conserved hypothe
7	492	37.9	233	2	S66100	conserved hypothe
8	396.5	30.5	262	2	E75516	conserved hypothe
9	389.5	30.0	274	2	H86937	conserved hypothe
10	382.5	29.4	272	2	A70955	hypothetical prote
11	359.5	27.7	246	2	D72320	conserved hypothe
12	337.5	26.0	273	2	D71326	conserved hypothe
13	189.5	14.6	262	2	F70165	conserved hypothe
14	175	13.5	592	2	H82031	probable biotin-[a
15	172	13.2	592	2	B81009	BirA protein/Bvg a
16	163.5	12.6	276	2	AI2292	hypothetical prote
17	160	12.3	248	2	H83111	hypothetical prote
18	147	11.3	257	2	S75559	hypothetical prote
19	140.5	10.8	242	2	A82637	conserved hypothe
20	134.5	10.4	267	2	I40327	baf protein - Bord
21	125.5	9.7	229	2	E70465	hypothetical prote
22	111	8.5	209	2	H81382	hypothetical prote
23	107.5	8.3	401	2	C83309	conserved hypothe
24	99	7.6	392	2	D70656	hypothetical prote
25	98	7.5	1536	2	A43855	high-molecular-we
26	97.5	7.5	456	2	JC4089	3alpha,7alpha,12a
27	97.5	7.5	5291	2	F90696	hypothetical prote
28	96.5	7.4	5188	2	B85547	probable RTX famil
29	96	7.4	460	2	AC2272	hypothetical prote

30	95.5	7.4	1682	1	C70588	probable mbtE prot
31	94.5	7.3	4735	2	T17463	rifamycin polyketi
32	94	7.2	530	2	AH3194	glycogen synthase
33	92.5	7.1	657	2	T25613	hypothetical prote
34	92.5	7.1	701	1	QOECNA	carbon starvation
35	92	7.1	384	2	S75916	probable soluble h
36	91.5	7.0	386	2	F72773	probable molybdopt
37	91.5	7.0	487	2	B32475	dnak-type molecula
38	91	7.0	327	2	AH2312	glycine cleavage T
39	91	7.0	464	2	H87208	cystathionine [bet
40	90.5	7.0	287	2	D90540	glucokinase (gluco
41	90	6.9	260	2	S36105	hgdc protein - Aci
42	90	6.9	478	2	T35759	pyruvate kinase -
43	90	6.9	636	2	AI0057	chaperone protein
44	89.5	6.9	701	2	E90708	carbon starvation
45	89.5	6.9	701	2	A85559	carbon starvation

ALIGNMENTS

RESULT 1  
B87489  
transcription activator, probable Baf family [imported] - Caulobacter crescentus  
C;Species: Caulobacter crescentus  
C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
C;Accession: B87489  
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; F  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser,  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A;Title: Complete Genome Sequence of Caulobacter crescentus.  
A;Reference number: A87249; MUID:21173698; PMID:11259647  
A;Accession: B87489  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-261 <STO>  
A;Cross-references: GB:AE005673; NID:gl3423392; PIDN:AAK23910.1; GSPDB:GN00148  
C;Genetics:  
A;Gene: CC1935  
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match	100.0%;	Score 1299;	DB 2;	Length 261;
Best Local Similarity	100.0%;	Pred. No. 3.7e-96;		
Matches 260;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MLLAIEQGNTNTMFAIHGASWVAQWRSAT	ESTRTADEYVVWLSQLLSMQGLGFRAIDAV	60
Db	2	MLLAIEQGNTNTMFAIHGASWVAQWRSAT	ESTRTADEYVVWLSQLLSMQGLGFRAIDAV	61
QY	61	IISVVVPQSIFNLRLNLSRRYFNVEPLVIGENAKLGIDVRIEKPSEACADRLVNAIGAAMV	120	
Db	62	IISVVVPQSIFNLRLNLSRRYFNVEPLVIGENAKLGIDVRIEKPSEACADRLVNAIGAAMV	121	
QY	121	YPGPLVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMQALHEAAKLPRIAIQRPAGNR	180	
Db	122	YPGPLVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMQALHEAAKLPRIAIQRPAGNR	181	
QY	181	IVGTDTVSAMQSGVFWGYISLIEGLVARIKAERGEPMTVIATGGVASLFEGATDSIDHFD	240	
Db	182	IVGTDTVSAMQSGVFWGYISLIEGLVARIKAERGEPMTVIATGGVASLFEGATDSIDHFD	241	
QY	241	SDLTIRGLLEIYRRNTIAES	260	
Db	242	SDLTIRGLLEIYRRNTIAES	261	

RESULT 2

F83660  
hypothetical protein BH0086 [imported] - Bacillus halodurans (strain C-125)  
C;Species: Bacillus halodurans  
C;Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C;Accession: F83660







E75516  
conserved hypothetical protein - Deinococcus radiodurans (strain R1)  
C;Species: Deinococcus radiodurans  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 20-Jun-2000  
C;Accession: E75516  
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A;Reference number: A75250; MUID:20036896; PMID:10567266  
A;Accession: E75516  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-262 <WHI>  
A;Cross-references: GB:AE001905; GB:AE000513; NID:g6458144; PIDN:AAF10040.1; PID:g645814  
A;Experimental source: strain R1  
C;Genetics:  
A;Gene: DR0461  
A;Map position: 1  
Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 30.5%; Score 396.5; DB 2; Length 262;  
Best Local Similarity 36.1%; Pred. No. 2.7e-24;  
Matches 92; Conservative 45; Mismatches 109; Indels 9; Gaps 5;  
QY 2 LLAIEQGNTNTMFAIHGASWVAQ-WRSATESTRTADEYVWVLSQLLSMOGLGFRAIDAV 60  
Db 6 LLAVDIGNTTTLGLADASGALHTWRTNRMLPDDLALQLHGLFTLAGAPIPR-AA 63  
QY 61 IISVVVPQSFIFNLRLNLSRRYFNVPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAMV 120  
Db 64 VLSSVAPPVGENYALAKRHFMDAFVSAENLPDVTVELDTPGSVGADRLCNLFGEAK- 122  
QY 121 YPGPL--VVIDSGTATTFDIVAADGAFEGGIIAPGINLSMQALHEAAAKLPRIAIPRA 177  
Db 123 YLGGLDYAVVDFGTSTNFDVVGRRFLGGILATGAQVSADALFARAALKLPRTLQAP- 181  
QY 178 GNRIVGTDVTSAMQSGVFWGYISLIEGLVARIKAEGERPMTVIATGGVASLFEGATDSID 237  
Db 182 -ETAIGKNTVHALQSLVFGYAEVMDGLLRRIRAEPLGEAVAVATGGFSRTVQICQEID 240  
QY 238 HFDSDLTIRGLLEIY 252  
Db 241 YYDETTLRGLVELW 255

RESULT 9  
H86937  
conserved hypothetical protein ML0232 [imported] - Mycobacterium leprae  
Species: Mycobacterium leprae  
C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
C;Accession: H86937  
R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho  
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,  
eam, M.A.; Rutherford, K.M.  
Nature 409, 1007-1011, 2001  
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq  
A;Title: Massive gene decay in the leprosy bacillus.  
A;Reference number: A86909; MUID:21128732; PMID:11234002  
A;Accession: H86937  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-274 <STO>  
A;Cross-references: GB:AL450380; NID:g13092576; PIDN:CAC29740.1; GSPDB:GN00147  
C;Genetics:  
A;Gene: ML0232  
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 30.0%; Score 389.5; DB 2; Length 274;  
Best Local Similarity 34.3%; Pred. No. 1e-23;  
Matches 93; Conservative 53; Mismatches 108; Indels 17; Gaps 5;

QY 1 MLLAIEQGNTNTMFAIHG----ASWVAQWRSATESTRTADEYVWVLSQLLSMOGLGFRA 56  
Db 1 MLLAIDVRNTHTVVGLLSGSKHEAKVVQWRIRTESEVTADELALITDGLI---GDDSER 57  
QY 57 IDAVIISVVVPQSFIFNLRLNLSRRYF-NVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAI 115  
Db 58 LAGAAALSTVPSVLHEVRIMLDQYWPSVPHVLIIEPGVRTGIPLLVDNPKVEVGADRIVNCL 117  
QY 116 GAAMVYPGPLVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMQALHEAAAKLPRIAIPRA 175  
Db 118 AAFHKFGQAAIIVDFGSSICVDVVSAGKEFLGGAIAPGVQVSSDAAAARSAAALRRVELAR 177  
QY 176 PAGNRIVGTDVTSAMQSGVFWGYISLIEGLVARIK-----AERGEPMTVIATGGVASL 228  
Db 178 P--RSVVGKNTVECMQAGVVFGFAGLVGDMRQDVVEEFGSLGNRVAVVATGHTAPL 235  
QY 229 FEGATDSIDHFDSDLTIRGLLEIYRRNTIAE 259  
Db 236 LLPELHTVDHYDRHLTLHGLRLVFERNREAQ 266

RESULT 10  
A70955  
hypothetical protein Rv3600c - Mycobacterium tuberculosis (strain H37RV)  
C;Species: Mycobacterium tuberculosis  
C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C;Accession: A70955  
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordo  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete gen  
A;Reference number: A70500; MUID:98295987; PMID:9634230  
A;Accession: A70955  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-272 <COL>  
A;Cross-references: GB:Z95557; GB:AL123456; NID:g3242276; PIDN:CAB08944.1; PID:g2113  
A;Experimental source: strain H37RV  
C;Genetics:  
A;Gene: Rv3600c  
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c  
Query Match 29.4%; Score 382.5; DB 2; Length 272;  
Best Local Similarity 34.0%; Pred. No. 3.7e-23;  
Matches 90; Conservative 53; Mismatches 107; Indels 15; Gaps 5;  
QY 1 MLLAIEQGNTNTMFAIHG----ASWVAQWRSATESTRTADEYVWVLSQLLSMOGLGFRA 56  
Db 1 MLLAIDVRNTHTVVGLLSGSKHEAKVVQWRIRTESEVTADELALITDGLI---GEDSER 57  
QY 57 IDAVIISVVVPQSFIFNLRLNLSRRYF-NVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAI 115  
Db 58 LTGTAALSTVPSVLHEVRIMLDQYWPSVPHVLIIEPGVRTGIPLLVDNPKVEVGADRIVNCL 117  
QY 116 GAAMVYPGPLVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMQALHEAAAKLPRIAIPRA 175  
Db 118 AAYDRFRKAAIIVDFGSSICVDVVSAGKEFLGGAIAPGVQVSSDAAAARSAAALRRVELAR 177  
QY 176 PAGNRIVGTDVTSAMQSGVFWGYISLIEGLVARIKAEGERPMTVIATGGVASLFE 230  
Db 178 P--RSVVGKNTVECMQAGAVFGFAGLVGDMRQDVVEEFGSLGNRVAVVATGHTAPLLL 235  
QY 231 GATDSIDHFDSDLTIRGLLEIYRRN 255  
Db 236 PELHTVDHYDQHLTLQGLRLVFERN 260

RESULT 11  
D72320  
conserved hypothetical protein - Thermotoga maritima (strain MSB8)  
C;Species: Thermotoga maritima



Search completed: June 24, 2003, 22:19:17  
Job time : 11.7262 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

-OM protein - protein search, using sw model

Run On: June 24, 2003, 21:51:40 ; Search time 5.23014 Seconds  
(without alignments)  
2061.866 Million cell updates/sec

Title: US-09-813-453A-51  
Perfect score: 1299  
Sequence: 1 MLLAIEQGNNTMFAIHGDA.....SDLTIRGLLEIYRRNTIAES 260

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues  
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	492	37.9	233	1 YACB_BACSU	P37564 bacillus su
2	134.5	10.4	267	1 BAF_BORPE	Q45338 bordetella
3	94	7.2	496	1 NIFE_RHILO	Q98ap4 rhizobium l
4	93	7.2	491	1 YC70_CORGL	P42531 corynebacte
5	92.5	7.1	657	1 HS7F_CAEEL	P11141 caenorhabdi
6	92.5	7.1	701	1 CSTA_ECOLI	P15078 escherichia
7	92	7.1	308	1 SPD1_DATST	Q96556 datura stra
8	90	6.9	260	1 HGDC_ACIFE	P11568 acidaminoco
9	90	6.9	496	1 NIFE_RHISN	P55673 rhizobium s
10	90	6.9	636	1 DNAK_HUMAN	Q8zim7 yersinia pe
11	89.5	6.9	478	1 CARL_HUMAN	Q9uhj6 homo sapien
12	89	6.9	556	1 CG48_SCHPO	P78750 schizosacch
13	88.5	6.8	469	1 LEU2_RHILO	Q98ef1 rhizobium l
14	88	6.8	399	1 PGK_RHIME	Q92m79 rhizobium m
15	88	6.8	1307	1 BCC4_ACEXY	Q9wx71 acetobacter
16	87.5	6.7	245	1 HIS4_BACSU	O35006 bacillus su
17	87.5	6.7	541	1 YQEB_ECOLI	Q46808 escherichia
18	87.5	6.7	555	1 YP85_MYCLE	Q49646 mycobacteri
19	87.5	6.7	576	1 MUTL_CHLMU	Q9pi95 chlamydia m
20	87.5	6.7	923	1 GYRA_PSEAE	P48372 pseudomonas
21	87	6.7	201	1 RISA_MYCTU	P71680 mycobacteri
22	87	6.7	637	1 DNAK_SALTI	Q829r1 salmonella
23	87	6.7	637	1 DNAK_SALTY	Q56073 salmonella
24	86.5	6.7	343	1 YJCR_ECOLI	P32716 escherichia
25	86.5	6.7	548	1 Y4VL_RHISN	Q53217 rhizobium s
26	86.5	6.7	1132	1 PHYL_PHYPA	P36505 physcomitre
27	86	6.6	537	1 CH61_MYCLE	P37578 mycobacteri
28	86	6.6	788	1 PUR2_YARLI	Q99148 y bifunctio
29	85	6.5	637	1 DNAK_ECOLI	P04475 escherichia
30	85	6.5	1706	1 CYAA_BORPE	P15318 bordetella
31	84.5	6.5	1705	1 CYAA_BORBR	Q57506 bordetella
32	83	6.4	465	1 LEU2_EC057	Q8xa00 escherichia
33	83	6.4	465	1 LEU2_ECOLI	P30127 escherichia

34	83	6.4	541	1	LCPT_BACSU	P55910 bacillus su
35	82.5	6.4	286	1	PTND_ECOLI	P08188 escherichia
36	82.5	6.4	469	1	LEU2_AGR75	Q8uby9 agrobacteri
37	82.5	6.4	476	1	LEU2_YERPE	Q8zih0 yersinia pe
38	82	6.3	609	1	COPA_PSESM	P12374 pseudomonas
39	81.5	6.3	434	1	ENO_STRPY	P82479 streptococc
40	81.5	6.3	769	1	VPI_BPPH6	P11126 bacterioph
41	81	6.2	363	1	MODC_RHOCA	Q08381 rhodobacter
42	81	6.2	487	1	IMDH_PASMU	Q916b7 pasteurella
43	81	6.2	637	1	DNAK_BRUME	Q8ye76 brucella me
44	81	6.2	644	1	DNAK_NITEU	O06430 nitrosomona
45	81	6.2	654	1	DNAK_CHLMU	P56836 chlamydia m

ALIGNMENTS

RESULT 1  
YACB\_BACSU  
ID YACB\_BACSU STANDARD; PRT; 233 AA.  
AC P37564;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein yacB.  
GN YACB.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=96051385; PubMed=7584024;  
RA Ogasawara N., Nakai S., Yoshikawa H.;  
RT "Systematic sequencing of the 180 kilobase region of the Bacillus  
subtilis chromosome containing the replication origin.";  
RL DNA Res. 1:1-14(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
Choi S.K., Codani J.J., Connerton I.F., Cumings N.J., Daniel R.A.,  
Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,  
Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
Joris E., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
Lee S.M., Levine A., Liu H., Masuda S., Mael C., Medigue C.,  
Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,  
Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,  
Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
subtilis.";  
RL Nature 390:249-256(1997).  
CC -I- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.

```
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D26185; BAA05305.1; -.
DR EMBL; Z99104; CABL11846.1; -.
DR Subtilist; BG10133; yacB.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 233 AA; 26217 MW; AAE96E732C15DF44 CRC64;

Query Match 37.9%; Score 492; DB 1; Length 233;
Best Local Similarity 45.2%; Pred. No. 2.3e-31;
Matches 95; Conservative 41; Mismatches 72; Indels 2; Gaps 1;

-y 1 MLLAIEQNTNTMFAIHGASWVAQWRSATESPTADEYVWLSQLLSMQGLGFRAIDAV 60
   ||| : |||| : : || : || : || : || : || : || : || : || : || :
Db 1 MLLVIDVGNNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
   ||||| : || : |||| : || : || : || : || : || : || : || : || : || :
QY 61 IISVVVPQSIFNLNLSRRYFNEPLVIGENAKLIGIDVRIEKPSEAGADRLVNAIGAMV 120
   ||||| : || : |||| : || : || : || : || : || : || : || : || : || :
Db 61 IISVVPPIMFALERMCTKYFHIEPIVGPGMKTGLNIKYDNPKEVGADRIVNAVAIHL 120
   ||||| : || : |||| : || : || : || : || : || : || : || : || : || :
QY 121 YPGPLVIDSGTATFDIVAADGAFEGGIIAPGINLSMQALHEAAKLPRIAIPAGNR 180
   ||| : |||| : : : |||| : || : |||| : || : |||| : || : |||| : || :
Db 121 YGNPLIVVDFGTATYCYIDENKQYMGGAIAPIGTISTEALYSRAAKLPRIETRP--DN 178
   || : |||| : || : |||| : || : |||| : || : |||| : || : |||| : || :
QY 181 IVGTDTVSAMQSGVFWGYISLIEGLVARIK 210
   || : |||| : || : |||| : || : |||| : || : |||| : || : |||| : || :
Db 179 IIGKNTVSAMQSGILFGYGVQVEGIVKRMK 208
   || : |||| : || : |||| : || : |||| : || : |||| : || : |||| : || :

RESULT 2
BAF_BORPE STANDARD; PRT; 267 AA.
AC Q45338; Q45373;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Bvg accessory factor.
GN BAF.
OS Bordetella pertussis.
: Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
: Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BP504;
RX MEDLINE=95325323; PubMed=7601846;
RA Deshazer D., Wood G.E., Friedman R.L.;
RT "Identification of a Bordetella pertussis regulatory factor required
RT for transcription of the pertussis toxin operon in Escherichia
RT coli.";
RL J. Bacteriol. 177:3801-3807(1995).
RN [2]
RP SEQUENCE OF 1-38 FROM N.A.
RC STRAIN-BP504;
RA Wood G.E., Friedman R.L.;
RT "Identification of a bira homolog in Bordetella pertussis.";
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 239-267 FROM N.A.
RC STRAIN-BP536;
RX MEDLINE=96419162; PubMed=8821935;
RA Allen A.G., Maskell D.J.;
RT "The identification, cloning and mutagenesis of a genetic locus
```

```
RT required for lipopolysaccharide biosynthesis in Bordetella
RT pertussis.";
RL Mol. Microbiol. 19:37-52(1996).
CC -!- FUNCTION: ACTIVATES TRANSCRIPTION OF THE PERTUSSIS TOXIN OPERON IN
CC A BVCAS-DEPENDENT MANNER. MAY INTERACT WITH THE ALPHA SUBUNIT OF
CC RNA POLYMERASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U12020; AAA75361.1; -.
DR EMBL; AF016461; AAC68834.1; -.
DR EMBL; X90711; CAA62242.1; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
KW Transcription regulation; Activator.
SQ SEQUENCE 267 AA; 27845 MW; 336A615F67B57901 CRC64;

Query Match 10.4%; Score 134.5; DB 1; Length 267;
Best Local Similarity 28.8%; Pred. No. 0.0014;
Matches 53; Conservative 18; Mismatches 78; Indels 35; Gaps 8;

QY 79 RYFNEPLVIGENAKLIGIDVRIEKPSEAGADRLVNAIGAMVYPG---PLVVIDSGTATT 135
   | : || : || : || : || : || : || : || : || : || : || : || : || :
Db 81 RWLRAQPLA-----MGLRNGYRNPDLGADRACWVGVLARQPSVHPPLLVASFGTATT 134
   | : || : || : || : || : || : || : || : || : || : || : || : || :
QY 136 FDIVAADGAFEGGIIAPGINLSMQALHEAAKLP---RIAQRPAQNRIVGTDTVSAMQS 192
   | : || : || : || : || : || : || : || : || : || : || : || : || :
Db 135 LDTIGPDNVFPGGLILPGPAMRGALAYGTAHLPLADGLVADYPIDTHQAIASGIAAAQA 194
   | : || : || : || : || : || : || : || : || : || : || : || : || :
QY 193 GVF---W-----GYISLIEGLVA-----RIKAERGEPTVIATGGVASLFEGATDSID 237
   | : || : || : || : || : || : || : || : || : || : || : || : || :
Db 195 GAIVRQWLAGRQRYGQAPETIYVAGGGWPEVRQEAER-----LLAVTGAAF---GATPQPT 246
   | : || : || : || : || : || : || : || : || : || : || : || : || :

238 HFDS 241
: ||
Db 247 YLDS 250

RESULT 3
NIFE_RHILO
ID NIFE_RHILO STANDARD; PRT; 496 AA.
AC Q98AP4;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Nitrogenase iron-molybdenum cofactor biosynthesis protein nife.
GN NIFE OR MLR5908.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
CC -!- FUNCTION: THIS PROTEIN MAY PLAY A ROLE IN THE BIOSYNTHESIS OF
CC THE PROSTHETIC GROUP OF NITROGENASE (FEMO COFACTOR).
CC -!- PATHWAY: Fe-Mo cofactor biosynthesis.
CC -!- SIMILARITY: BELONGS TO THE NIFD/NIFK/NIFE/NIFN FAMILY.
```



CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; U88315; AAB42371.1; -;  
DR EMBL; X07678; CAA30525.1; -;  
DR PIR; S03210; S03210.  
DR PIR; B32475; B32475.  
DR HSSP; P04475; 1DG4.  
DR WormPep; C37H5.8; CE08631.  
DR InterPro; IPR001023; Hsp70.  
DR Pfam; PF00012; HSP70; 1.  
DR PRINTS; PR00301; HEATSHOCK70.  
DR ProDom; PD000089; Hsp70; 1.  
DR PROSITE; PS00297; HSP70\_1; 1.  
DR PROSITE; PS00329; HSP70\_2; 1.  
DR PROSITE; PS01036; HSP70\_3; 1.  
KW ATP-binding; Heat shock; Multigene family; Mitochondrion;  
KW Transit peptide.  
FT TRANSIT 1 27 MITOCHONDRION (POTENTIAL).  
CH 28 657 HEAT SHOCK 70 KDA PROTEIN F.  
CONFLICT 138 138 V -> E (IN REF. 2).  
CONFLICT 140 140 S -> P (IN REF. 2).  
CONFLICT 167 167 A -> P (IN REF. 2).  
CONFLICT 347 348 KA -> NV (IN REF. 2).  
SEQUENCE 657 AA; 70844 MW; 637A1636B6ABB836 CRC64;

Query Match 7.1%; Score 92.5; DB 1; Length 657;  
Best Local Similarity 21.9%; Pred. No. 7.1;  
Matches 57; Conservative 44; Mismatches 102; Indels 57; Gaps 13;

QY 2 LLAIEQNTNMFATHDGASWVAQWRSATSTRTADEYVWLSQLLSMOGLGFRDAVI 61  
Db 33 VIGIDLTGNTSCVSIMEGKT--PKVIENAEGRVTPSTVFTAD-----GERLVGAPA 83  
QY 62 ISSVVPQSIFNL----RNLSRRYFNVE-----PLVIGENAKLGIDV--RIEKPSE 105  
Db 84 KRQAVTNSANTLFTATKRLIGRRYEDPEVQKDLKVVPYKIVKASNGDAWVEAQGVYSPSQ 143  
QY 106 AGADRLVNAIGAAMVYPGLVVIDSGTAT--FDIVAADGAFEGGIIAPGINLSMQALH 162  
Db 144 VGAFVLMKMKETAESYLG--TTVNNAVTVPAYFNDSQRQATKDAGQIS-GLNV-LRVIN 199  
QY 163 EAAAKLPRIATQRPAGNRIV-----GTDTVSAM--QSGVF-----WG Y 198  
Db 200 EPTAAALAYGLDKDAGDKIIAVYDLGGTFDVSILEIQGVFEVKSTNGDTFLGGEDFDH 259

RESULT 6  
CSTA\_ECOLI  
ID CSTA\_ECOLI STANDARD; PRT; 701 AA.  
AC P15078; P23517; P77740;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Carbon starvation protein A.  
GN CSTA OR B0598.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12;  
RX MEDLINE=91162636; PubMed=1848300;  
RA Schultz J.E., Matin A.;  
RT "Molecular and functional characterization of a carbon starvation  
RL J. Mol. Biol. 218:129-140(1991).

[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,  
Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D.,  
Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;  
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE OF 1-200 FROM N.A.  
RX MEDLINE=89123155; PubMed=2521622;  
RA Liu J., Duncan K., Walsh C.T.;  
RT "Nucleotide sequence of a cluster of Escherichia coli enterobactin  
biosynthesis genes: identification of entA and purification of its  
product 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase.";  
RL J. Bacteriol. 171:791-798(1989).  
CC -1- FUNCTION: PEPTIDE UTILIZATION DURING CARBON STARVATION.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
CC (Potential).  
CC -1- INDUCTION: By carbon starvation.  
CC -1- SIMILARITY: BELONGS TO THE CSTA FAMILY.  
CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A  
CC FRAMESHIFT IN POSITION 543.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X52904; CAA37086.1; ALT\_FRAME.  
DR EMBL; X52904; CAA37087.1; ALT\_FRAME.  
DR EMBL; AE000165; AAC73699.1; -;  
DR EMBL; U82598; AAB40798.1; -;  
DR EMBL; M24148; -; NOT\_ANNOTATED\_CDS.  
DR EMBL; D90701; BAA35227.1; -;  
DR PIR; E32047; QOECNA.  
DR PIR; S14411; S14411.  
DR PIR; S14412; S14412.  
DR EcoGene; EG10167; csta.  
DR InterPro; IPR003706; Csta.  
DR Pfam; PF02554; Csta; 1.  
KW Transmembrane; Inner membrane; Complete proteome.  
FT TRANSMEM 7 27 POTENTIAL.  
FT TRANSMEM 34 54 POTENTIAL.  
FT TRANSMEM 87 107 POTENTIAL.  
FT TRANSMEM 118 138 POTENTIAL.  
FT TRANSMEM 162 182 POTENTIAL.  
FT TRANSMEM 190 210 POTENTIAL.  
FT TRANSMEM 218 238 POTENTIAL.  
FT TRANSMEM 256 276 POTENTIAL.  
FT TRANSMEM 281 301 POTENTIAL.  
FT TRANSMEM 325 345 POTENTIAL.  
FT TRANSMEM 373 393 POTENTIAL.  
FT TRANSMEM 396 416 POTENTIAL.  
FT TRANSMEM 440 460 POTENTIAL.  
FT TRANSMEM 464 484 POTENTIAL.  
FT TRANSMEM 524 544 POTENTIAL.  
FT TRANSMEM 551 571 POTENTIAL.  
FT TRANSMEM 578 598 POTENTIAL.  
FT TRANSMEM 644 664 POTENTIAL.  
FT CONFLICT 129 129 A -> G (IN REF. 1).  
FT SEQUENCE 701 AA; 75105 MW; 2DA85B2D96FD19E5 CRC64;



Query Match 7.1%; Score 92.5; DB 1; Length 701;  
Best Local Similarity 22.2%; Pred. No. 7.6;  
Matches 57; Conservative 37; Mismatches 92; Indels 71; Gaps 13;

\*QY 13 MFAIHDGASWVAQ---WRSATESTRFADEYVWVLSQLLSMQGLGFRAIDAVIISVV PQS 69  
Db 230 IFAIISG-GWVAESPTWAPYDFGTG---QLTWM-----LVGYG-----VAAVLPVW 273  
QY 70 IFNLRLNRRYFNVEPLVIGENAKLGIDVRIEKP--SEAGADRLVNAIG---AAMVYGP 124  
Db 274 LL-----LAPRDYLSFLKIGTIVGLAVGILIMRPTLTMPALTKFVDGTGPVWTGNLFPFL 329  
QY 125 LVVIDSGTATTFDIVAADG-----AFEGG-----IIAPGI 154  
Db 330 FITIACGAVSGFHALISSGTTTPKMLANEGQACFYGGMLMESFVAINALVSACIIDPGV 389  
QY 155 NLSMQALHEAAKLPRIAIOQPAGNRIVGTDTVTSAMQSGVFWGYISLIEGLVARIKAE 214  
b 390 YFAMNS-----PMAVLAPAGTADV---VASAAQVVSSWGF-SITPDTLNQIASEVG 436  
QY 215 EPMTVIATGGVVASLFEG 231  
Db 437 EQSIISRAGGAPTAVG 453

RESULT 7  
SPDI\_DATST STANDARD; PRT; 308 AA.  
AC Q96556;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Spermidine synthase 1 (EC 2.5.1.16) (Putrescine aminopropyltransferase 1) (SPDSY 1).  
OS Datura stramonium (Jimsonweed) (Common thornapple).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids I; Solanales; Solanaceae; Datura.  
OX NCBI\_TaxID=4076;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=D5/15; TISSUE=Root;  
RA Michael A.J.;  
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- CATALYTIC ACTIVITY: S-adenosylmethionine + putrescine = 5'-methylthioadenosine + spermidine.  
CC -1- PATHWAY: Spermidine biosynthesis; fifth (last) step.  
CC -1- SIMILARITY: BELONGS TO THE SPERMIDINE/SPERMINE SYNTHASES FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Y08252; CAA69420.1; -.  
DR InterPro; IPR000051; SAM\_bind.  
DR InterPro; IPR001045; Sprimine\_synthase.  
DR Pfam; PF01564; Spermine\_synth; 1.  
DR TIGRFAMs; TIGR00417; speE; 1.  
DR PROSITE; PS01330; SPERMIDINE\_SYNTHASE; 1.  
KW Spermidine biosynthesis; Transferrase; Multigene family.  
FT DOMAIN 94 135 BINDING TO DECARBOXYLATED SAM (POTENTIAL).  
FT SEQUENCE 308 AA; 34033 MW; 820631DC791EF0DB CRC64;

Query Match 7.1%; Score 92; DB 1; Length 308;  
Best Local Similarity 25.5%; Pred. No. 3.2;  
Matches 52; Conservative 27; Mismatches 75; Indels 50; Gaps 9;

QY 73 LRNLRRYFNVEPLVIGENAKLGIDVRIE-KPSEAGADRLVNAIGAAMVYPGPLVVIDSG 131  
Db 107 LREVS-RHSSVEQIDICEIDKMVIDSVKQFFPN-----VAIGYEDPRVKLHV 153  
QY 132 TATTFDIVAADGAFEGGI-----IAPGINLSMQALHEAAKLPRIAIOQPAGNRIVGTD 185  
Db 154 DGVAFLKFVAEGTYDAVIDVSSDPIGPAQELFEKPFESVAR-----ALRPGG----- 201  
QY 186 TVSAMQSGVFWGYISLIEGLVARIK-AERGE-----PMTVI-----ATGGVASL 228  
Db 202 -VVCTQAESIWLHMHIEDIVANCRQIFKGSVNYAWTTVTPYPSGVIGFMLCSTEGPAVD 260  
QY 229 FEGATDSIDHFDSDLTIRGLLEIY 252  
Db 261 FKNPINVDADDSHTKTRGPKFY 284

RESULT 8  
HGDC\_ACIFE STANDARD; PRT; 260 AA.  
AC P11568; Q44042;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Activator of (R)-2-hydroxyglutaryl-CoA dehydratase (2-hydroxyglutaryl-CoA dehydratase component A).  
DE HGDC.  
GN Acidaminococcus fermentans.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Acidaminococcaceae;  
OC Acidaminococcus.  
OX NCBI\_TaxID=905;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 25085;  
RX MEDLINE=93374040; PubMed=8365476;  
RA Bendrat K., Mueller U., Klees A.-G., Buckel W.;  
RT "Identification of the gene encoding the activator of (R)-2-hydroxyglutaryl-CoA dehydratase from Acidaminococcus fermentans by gene expression in Escherichia coli.";  
RL FEBS Lett. 329:329-331(1993).  
RN [2]  
RP SEQUENCE OF 129-260 FROM N.A.  
RC STRAIN=ATCC 25085;  
RX MEDLINE=89276363; PubMed=26593350;  
RA Dutschko R., Wohlfarth G., Buckel P., Buckel W.;  
RT "Cloning and sequencing of the genes of 2-hydroxyglutaryl-CoA dehydratase from Acidaminococcus fermentans.";  
RL Eur. J. Biochem. 181:741-746(1989).  
RN [3]  
RP CHARACTERIZATION.  
RC STRAIN=ATCC 25085;  
RX MEDLINE=95331308; PubMed=7607244;  
RA Mueller U., Buckel W.;  
RT "Activation of (R)-2-hydroxyglutaryl-CoA dehydratase from Acidaminococcus fermentans.";  
RL Eur. J. Biochem. 230:698-704(1995).  
RN [4]  
RP CHARACTERIZATION.  
RX MEDLINE=20558245; PubMed=11106419;  
RA Hans M., Buckel W., Bill E.;  
RT "The iron-sulfur clusters in 2-hydroxyglutaryl-CoA dehydratase from Acidaminococcus fermentans. Biochemical and spectroscopic investigations.";  
RL Eur. J. Biochem. 267:7082-7093(2000).  
RN [5]  
RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).  
RX MEDLINE=21140179; PubMed=11243821;  
RA Locher K.P., Hans M., Yeh A.P., Schmid B., Buckel W., Rees D.C.;  
RT "Crystal structure of the Acidaminococcus fermentans 2-hydroxyglutaryl-CoA dehydratase component A.";  
RL J. Mol. Biol. 307:297-308(2001).  
CC -1- FUNCTION: REQUIRED FOR THE ACTIVATION OF (R)-2-HYDROXYGLUTARYL-COA DEHYDRATASE. THIS PROTEIN IS EXTREMELY SENSITIVE TOWARDS OXYGEN.

CC -!- COFACTOR: BINDS ONE 4FE-4S CLUSTER PER DIMER.  
CC -!- SUBUNIT: HOMODIMER.  
CC -!- SIMILARITY: TO E.COLI YJIL AND M.JANNASCHII MJ0004 AND MJ0800.  
CC  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC  
CC EMBL; X59645; CAA42196.1; -  
DR EMBL; X14252; CAA32464.1; -  
DR PIR; S04476; S04476.  
DR PDB; 1HUX; 21-MAR-01.  
DR InterPro; IPR002731; ATPase\_BadF.  
DR Pfam; PF01869; BcrAD\_BadFG; 1.  
DR TIGRFAMS; TIGR00241; CoA\_E-activ; 1.  
DR ProDom; PD006344; ATPase\_BadF; 1.  
DR Iron-sulfur; 4Fe-4S; 3D-structure.  
DR METAL 127 127 IRON-SULFUR (4FE-4S).  
DR METAL 166 166 IRON-SULFUR (4FE-4S).  
FT CONFLICT 196 196 A -> P (IN REF. 2).  
FT CONFLICT 209 209 V -> L (IN REF. 2).  
FT CONFLICT 214 224 MTGGVAQNYGV -> HDRRCSPELWL (IN REF. 2).  
SQ SEQUENCE 260 AA; 27269 MW; 7E97044DBE805AC CRC64;

Query Match 6.9%; Score 90; DB 1; Length 260;  
Best Local Similarity 25.4%; Pred. No. 3.8;  
Matches 48; Conservative 29; Mismatches 58; Indels 54; Gaps 11;  
QY 89 GENAKLGI-DVRIEKPSEAGADRLVNAIGAAMVYPGLVVID-----SGTATT 135  
Db 68 GRNSLEGADKQMSLS-----CHAMGASFIWPNVHTVIDIGGDVKVIHVGWTWN 120  
QY 136 FDI----VAADGAFEGGIAPGINLSMQALHEAAK-LPRAIQRP-----AGNRIV---- 182  
Db 121 FQMDKCAAGTGRFL-DVMANILEVKVSDLAELGAKSTKRVASISSTCTVFAESEVISQLS 179  
QY 183 -GTDVTSAMQSGVFWGVIISLIEGLVARIKAEGERPMTVIATGGVASLFEGATDSIDHFD 241  
Db 180 KGTDKIDII-AGIHRVASRVIGLANRV-----GIVKDVVMTGGVA-----Q 220  
QY 242 DLTIRGLLE 250  
Db 221 NYGVGALE 229

RESULT 9  
ID NIFE\_RHISN STANDARD; PRT; 496 AA.  
AC P55673;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Nitrogenase iron-molybdenum cofactor biosynthesis protein nife.  
GN NIFE OR Y4VN.  
OS Rhizobium sp. (strain NGR234).  
OG Plasmid sym pNGR234a.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Rhizobium.  
OX NCBI\_TaxID=394;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-97305956; PubMed-9163424;  
RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,  
RA Perret X.;  
RT "Molecular basis of symbiosis between Rhizobium and legumes."  
RL Nature 387:394-401(1997).  
CC -!- FUNCTION: THIS PROTEIN MAY PLAY A ROLE IN THE BIOSYNTHESIS OF  
CC THE PROSTHETIC GROUP OF NITROGENASE (FMO COFACTOR).  
CC -!- PATHWAY: Fe-Mo cofactor biosynthesis.

CC -!- SIMILARITY: BELONGS TO THE NIFD/NIFK/NIFE/NIFN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC  
CC EMBL; AE000102; AAB91902.1; -  
DR HSSP; P07328; 3MIN.  
DR InterPro; IPR000318; Nitrognse\_compl.  
DR InterPro; IPR000510; Oxred\_nitrognse1.  
DR Pfam; PF00148; oxidored\_nitro; 1.  
DR TIGRFAMS; TIGR01283; nife; 1.  
DR PROSITE; PS00699; NITROGENASE\_1\_1; 1.  
DR PROSITE; PS00090; NITROGENASE\_1\_2; 1.  
KW Nitrogen fixation; Plasmid.  
SQ SEQUENCE 496 AA; 54793 MW; D78472D8F5410A3A CRC64;

Query Match 6.9%; Score 90; DB 1; Length 496;  
Best Local Similarity 25.4%; Pred. No. 8;  
Matches 46; Conservative 30; Mismatches 67; Indels 38; Gaps 8;  
QY 61 IISVVVPQ-----SIFNLRLNLSRRYFNVFVIGENAKLGIDVRIEKPSE-----AG 107  
Db 183 VIGTVEPDDAGPCDINILGEFNLGSEFWLVKPLL-----ERLGI RVRACIPGDARYLDIAS 238  
QY 108 ADRLVNAIGAAMVYPGLVVIDSGTATTFDVAADGAFEGGIAPGINLSMQALHEAAK 167  
Db 239 AHR---ARAAMLVCSTALINLARKMEERWDIPFEGSYG-----ITATSEALRQ 285  
QY 168 LPRAIQRPAGNRIVG-TDVTSAMQSGVFWGVIISLIEGLVARIKAEGERPMTVIATGGVA 226  
Db 286 IADLLVKKGTDLLEILDRTDALIAEEEAIAW---KKLEEYRPLKGR-----VLINTGGVK 338  
QY 227 S 227  
Db 339 S 339

RESULT 10  
ID DNAK\_YERPE STANDARD; PRT; 636 AA.  
AC Q8ZIM7;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Chaperone protein dnaK (Heat shock protein 70) (Heat shock 70 kDa  
protein) (HSP70).  
GN DNAK OR YPO0468.  
OS Yersinia pestis.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Yersinia.  
OX NCBI\_TaxID=632;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CO-92 / Biovar Orientalis;  
RX MEDLINE-21470413; PubMed-11586360;  
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,  
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,  
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,  
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,  
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,  
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;  
RT "Genome sequence of Yersinia pestis, the causative agent of plague."  
RL Nature 413:523-527(2001).  
CC -!- FUNCTION: Acts as a chaperone (By similarity).  
CC -!- INDUCTION: BY STRESS CONDITIONS E.G. HEAT SHOCK (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.





RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
RT "The genome sequence of Schizosaccharomyces pombe";  
RL Nature 415:871-880(2002).  
RN [2]  
RP SEQUENCE OF 269-556 FROM N.A.  
RC STRAIN=PR745;  
RX MEDLINE=98162722; PubMed=9501991;  
RA Yoshioka S., Kato K., Nakai K., Okayama H., Nojima H.;  
RT "Identification of open reading frames in Schizosaccharomyces pombe  
RT CDNAS";  
RL DNA Res. 4:363-369(1997).  
CC -1- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).  
CC -1- SIMILARITY: BELONGS TO THE CGI-48 FAMILY OF WD-REPEAT PROTEINS.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; AL022299; CAA18383.1; -  
CC EMBL; D89098; BAA13761.1; -  
CC InterPro; IPR001680; WD40.  
CC Pfam; PF00400; WD40; 4.  
CC SMART; SM00320; WD40; 4.  
DR PROSITE; PS00678; WD\_REPEATS\_1; FALSE\_NEG.  
DR PROSITE; PS50082; WD\_REPEATS\_2; 1.  
DR PROSITE; PS50294; WD\_REPEATS\_REGION; 1.  
KW Hypothetical protein; Repeat; WD repeat.  
FT REPEAT 216 254 WD 1.  
FT REPEAT 259 298 WD 2.  
FT REPEAT 390 429 WD 3.  
FT REPEAT 438 479 WD 4.  
FT REPEAT 485 523 WD 5.  
SQ SEQUENCE 556 AA; 62715 MW; 2DF02D921F9D5D7D CRC64;  
  
Query Match 6.9%; Score 89; DB 1; Length 556;  
Best Local Similarity 21.5%; Pred. No. 11;  
Matches 52; Conservative 35; Mismatches 101; Indels 54; Gaps 9;  
  
Db 23 VAQWRSATESRTADEYVWVLSQLLSMOGLGFRAIDAVIISVVPQSFNLRNLSRRYFN 82  
150 VPEWAKKQDVTEEDDEF-----NALSEKSVIPKSLKSLFKSSVSYIN 191  
83 VEPLVIGE---NAKLIGIDVRIEKPSEAGADRLVNAIGAAMVYPGVLVVIDSGTATFDIV 139  
192 QSSKLLAPGTINIKRLKDANFQAPSHSG-----IRCMSIHPYFPLLLTCGFDRTLRIY 244  
140 AADGAFEGGIIAPGINLSMQALHEAAKLPRIATQRPAGNRIVGTDTVSAMQSGVFWGYI 199  
245 QLDGK-----VNPLVTSLHLRSSAL-QTALFHPDGKRVIAAGRRKYM---YIWDLE 291  
200 SLIEGLVARIKAERG-----EPMTVIATGGVASLFEAGTDSID-----HFDSDLTIRG 247  
292 SAQVQKVSRYMQENFQPSMERFHVDPDTGKYIAL-EGRSGHINLLHALTGQFATSEKIEG 350  
248 LL 249  
351 VL 352  
  
RESULT 13  
LEU2\_RHILO

ID LEU2\_RHILO STANDARD; PRT; 469 AA.  
AC Q98EF1;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE 3-isopropylmalate dehydratase large subunit (EC 4.2.1.33)  
DE (isopropylmalate isomerase) (Alpha-IPM isomerase) (IPMI).  
GN LEUC OR ML4272.  
OS Rhizobium loti (Mesorhizobium loti).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Phyllobacteriaceae; Mesorhizobium.  
OX NCBI\_TaxID=381;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAFF303099;  
RX MEDLINE=21082930; PubMed=11214968;  
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,  
RA Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
RT Mesorhizobium loti";  
RL DNA Res. 7:331-338(2000).  
CC -1- FUNCTION: Catalyzes the isomerization between 2-isopropylmalate  
CC and 3-isopropylmalate, via the formation of 2-isopropylmaleate.  
CC -1- CATALYTIC ACTIVITY: 3-isopropylmalate = 2-isopropylmaleate +  
CC H(2)O.  
CC -1- CATALYTIC ACTIVITY: 2-isopropylmaleate + H(2)O = 2-  
CC isopropylmalate.  
CC -1- PATHWAY: Leucine biosynthesis; second step.  
CC -1- SUBUNIT: Heterodimer of leuc and leud (By similarity).  
CC -1- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY. LEUC 1  
CC SUBFAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; AP003003; BAB50968.1; -  
CC InterPro; IPR001030; Aconitase\_N.  
CC InterPro; IPR004430; Leuc.  
CC Pfam; PF00330; aconitase; 1.  
CC PRINTS; PR00415; ACONITASE.  
CC ProDom; PD000511; Aconitase\_N; 1.  
CC TIGRFAMs; TIGR00170; leuc; 1.  
CC PROSITE; PS00450; ACONITASE\_1; 1.  
CC PROSITE; PS01244; ACONITASE\_2; 1.  
KW Leucine biosynthesis; lyase; Iron-sulfur; 4Fe-4S; Complete proteome.  
FT METAL 350 350 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
FT METAL 410 410 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
FT METAL 413 413 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
SQ SEQUENCE 469 AA; 50871 MW; 641532D3D4F06888 CRC64;  
  
Query Match 6.8%; Score 88.5; DB 1; Length 469;  
Best Local Similarity 27.8%; Pred. No. 9.8;  
Matches 54; Conservative 26; Mismatches 69; Indels 45; Gaps 13;  
  
QY 64 SVVPQSIFNLRNLSRRYFNVEPLVIGENAK-LGIDVRIEKPSEAGADRLVNAIG--AAMV 120  
Db 68 STSPERKFGIKNEESR---IQVEALAKNAKDFGVVEYSENDIRQG---IVHIIGPEQGT 121  
QY 121 YPGPLVVI-DSGTATTFDIVAADGAFEGGIIAPGINLSMQALHEAAKLPRIATQRPAGN 179  
Db 122 LPGMTIVCGDSHTST-----HGAF--GALAHGIGTS-EVEHVLAT---QTLIQKAKN 168  
QY 180 RIVGTDTVSAMQSGVFWGYISLIEGLVAR--IKAEGERPMTVIATGGVASLFEAGTDSID 237  
Db 169 MLVRVDG-----QLPEGVTAKDIILAIIGE---IGTAGGTGYVIEYAGEAI- 211



```
QY 238 HFDSDLTIRGLLEI 251
Db 212 ---RSLMEGRMTI 222

RESULT 14
PGK_RHIME
ID PGK_RHIME STANDARD; PRT; 399 AA.
AC Q92M79;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Phosphoglycerate kinase (EC 2.7.2.3).
GN PGK OR R02765 OR SMC03981.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
NCBI_TaxID=382;
X N
N [1]
SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
CC -!- CATALYTIC ACTIVITY: ATP + 3-phospho-D-glycerate - ADP + 3-
CC phospho-D-glyceroyl phosphate.
CC -!- PATHWAY: Second phase of glycolysis; second step.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL591791; CAC47344.1; -
R InterPro; IPR001576; PGK.
R Pfam; PF00162; PGK; 1.
DR PROSITE; PS00111; PGLYCERATE_KINASE; 1.
KW Transferase; Kinase; GLYCOLYSIS; Complete proteome.
SQ SEQUENCE 399 AA; 41909 MW; 968386333CF000126 CRC64;

Query Match 6.8%; Score 88; DB 1; Length 399;
Best Local Similarity 25.0%; Pred. No. 8.9;
Matches 58; Conservative 25; Mismatches 65; Indels 84; Gaps 9;

QY 50 QGLG--FRAIDAVISSVWPQSIFNLRNLSRRYFNVEPLVIG----- 89
Db 181 KGLGNPKRPVVAIVGGAKVSTKIDLLQNLVKK---VDALVIGGGMANTFLAAQGVGVKS 237
QY 90 -----ENAK-----LGDVRIEKPSEAGADRLVNAIGAAMVYPGPLVVI 128
Db 238 LCEHDLAETAKAILAAASEAGCAIVLPVDGVVAREFKAGADNEVVDIKA---IPADAMVL 294
QY 129 DSG-----TATTFDIVAADGAFEGGIAPGINLSMQALHEAAKLPRIAIQRP 176
Db 295 DVGPKSIEAINEWISRAETLVWNGPLGAFE---IAPFDRATVAARKHAAARTRAGSLVSV 351
QY 177 AGNRIVGTDTVSAMQSGVFWGYISLIEGLVARIKAEGERPMTVIATGGVASL 228
Db 352 AG---GGDTVAALN-----HAEVADFTYVSTAGGAF 381
```

```
RESULT 15
BCC4_ACEXY
ID BCC4_ACEXY STANDARD; PRT; 1307 AA.
AC Q9WX71;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cellulose synthase 2 operon protein C precursor.
GN BCSC11.
OS Acetobacter xylinus.
OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
OC Gluconacetobacter.
OX NCBI_TaxID=28448;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 7664 / IFO 13693;
RX PubMed=10382968;
RA Umeda Y., Hirano A., Ishibashi M., Akiyama H., Onizuka T., Ikeuchi M.,
RA Inoue Y.;
RT "Cloning of cellulose synthase genes from Acetobacter xylinum JCM
RT 7664: implication of a novel set of cellulose synthase genes.";
RL DNA Res. 6:109-115(1999).
CC -!- FUNCTION: Required for maximal bacterial cellulose synthesis (By
CC similarity).
CC -!- PATHWAY: Bacterial cellulose biosynthesis.
CC -!- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC -!- SIMILARITY: BELONGS TO THE ACSC/BCSC FAMILY.
CC -!- SIMILARITY: CONTAINS 9 TPR REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB015803; BAA77596.1; -
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 5.
DR SMART; SM00028; TPR; 3.
KW Cellulose biosynthesis; Repeat; TPR repeat; Membrane; Outer membrane;
KW Signal.
FT SIGNAL 1 55 POTENTIAL.
FT CHAIN 56 1307 CELLULOSE SYNTHASE 2 OPERON PROTEIN C.
FT REPEAT 97 130 TPR 1.
FT REPEAT 270 303 TPR 2.
FT REPEAT 339 372 TPR 3.
FT REPEAT 374 406 TPR 4.
FT REPEAT 458 491 TPR 5.
FT REPEAT 493 525 TPR 6.
FT REPEAT 528 561 TPR 7.
FT REPEAT 754 787 TPR 8.
FT REPEAT 788 821 TPR 9.
SQ SEQUENCE 1307 AA; 141286 MW; 6FE5A5F38AD9FE53 CRC64;

Query Match 6.8%; Score 88; DB 1; Length 1307;
Best Local Similarity 23.2%; Pred. No. 35;
Matches 55; Conservative 40; Mismatches 96; Indels 46; Gaps 10;

QY 53 GFRAIDAVII---SSVVPQSIFNLRNLSRRYFNVEPLVIGENAKLIGDVRIEKPSEAGAD 109
Db 934 GAQAMDRQIVYLRDSVSPQ--FDANTFVRSRTGVAGL--GQLTEFAVPITATLPFESWDH 989
QY 110 RLVNAIGAAMVYPG-PLVVIDSGTATTFDIVAADGAFEGG---IAPGINLSMQALH--- 162
Db 990 RLSEFSVTPTLLFTGDPLT--NAVSAHQFGTAVAVNGARPWGYHHYTTQGVGLSLNVNRWF 1047
QY 163 -----EAAAKLPRIAIQRPAGNRIVGTDTVSAMQ-----SGVFW 196
Db 1048 AADVGSPLGFPITNVVGGLEFAPRLTRNLGLRISGGRMVTDSLSYAGERDPGTGKLW 1107
```

OY 197 GYISLIEGLVARIKAEGERPMTVIATGGVASLFEGATDSIDHFDSDLTIRGLLEIYR 253  
| : : | | : | | | | | : | : : : | : : :  
Db 1108 GGVTRLFGHGALEWSARG--WNAYAGGGFAYL--GGTNVIGNTETEAGAGGGSATVWQ 1160

Search completed: June 24, 2003, 22:11:50  
Job time : 6.23014 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:59:48 ; Search time 22.8708 Seconds  
(without alignments)  
2342.388 Million cell updates/sec

Title: US-09-813-453A-51  
Perfect score: 1299  
Sequence: 1 MLLAIEQGNTNTMFAIHGDA.....SDLTIRGLLEIYRRNTIAES 260

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues 671580  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_21:  
1: sp\_archaea:  
2: sp\_bacteria:  
3: sp\_fungi:  
4: sp\_human:  
5: sp\_invertebrate:  
6: sp\_mammal:  
7: sp\_mhc:  
8: sp\_organelle:  
9: sp\_phage:  
10: sp\_plant:  
11: sp\_rodent:  
12: sp\_virus:  
13: sp\_vertebrate:  
14: sp\_unclassified:  
15: sp\_rvirus:  
16: sp\_bacteriap:  
17: sp\_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1299	100.0	261	16 Q9A6Z1	Q9a6z1 caulobacter
2	598	46.0	254	16 Q9KGH5	Q9kgh5 bacillus ha
3	568	43.7	258	2 Q9F985	Q9f985 bacillus st
4	551	42.4	259	16 Q8XHL5	Q8xhl5 clostridium
5	530	40.8	255	16 Q8R7M2	Q8r7m2 thermoanaer
6	523	40.3	259	16 Q8YAC5	Q8yac5 listeria mo
7	519	40.0	273	16 Q97EB4	Q97eb4 clostridium
8	517	39.8	265	16 Q9X8N6	Q9x8n6 streptomyce
9	516	39.7	259	16 Q92F54	Q92f54 listeria in
10	423	32.6	256	16 Q8RFE4	Q8rfe4 fusobacteri
11	396.5	30.5	262	16 Q9RX54	Q9rx54 deinococcus
12	389.5	30.0	274	16 Q9CD56	Q9cd56 mycobacteri
13	382.5	29.4	272	16 O06282	O06282 mycobacteri
14	359.5	27.7	246	16 Q9WZY5	Q9wzy5 thermotoga
15	337.5	26.0	273	16 O83446	O83446 treponema p
16	307.5	23.7	212	2 O32514	O32514 desulfovibr

17	201	15.5	295	16 Q8Y2M4	Q8y2m4 ralstonia s
18	189.5	14.6	262	16 O51477	O51477 borrelia bu
19	175	13.5	592	16 Q9JWI7	Q9jwi7 neisseria m
20	172	13.2	592	16 Q9JXF1	Q9jxf1 neisseria m
21	163.5	12.6	276	16 Q8YQD7	Q8yqd7 anabaena sp
22	160	12.3	248	16 Q9HWC1	Q9hwc1 pseudomonas
23	147	11.3	257	16 P74045	P74045 synechocyst
24	140.5	10.8	242	16 Q9PCI4	Q9pci4 xylella fas
25	125.5	9.7	229	16 O67753	O67753 aquifex ao
26	111	8.5	209	16 Q9PIA9	Q9pia9 campylobact
27	107.5	8.3	401	16 Q9IOF1	Q9iof1 pseudomonas
28	102.5	7.9	1869	16 Q92JV3	Q92jv3 rhizobium m
29	99.5	7.7	729	2 P96168	P96168 vibrio harv
30	99.5	7.7	980	10 Q949G8	Q949g8 malus flori
31	99	7.6	392	16 P96214	P96214 mycobacteri
32	98	7.5	1536	2 Q48031	Q48031 haemophilus
33	97.5	7.5	456	2 Q54101	Q54101 saccharopol
34	97.5	7.5	5291	16 Q8X2T1	Q8x2t1 escherichia
35	96.5	7.4	5188	16 Q8X4H5	Q8x4h5 escherichia
36	96	7.4	460	16 Q8YQT7	Q8yqt7 anabaena sp
37	95.5	7.4	1787	16 O86329	O86329 mycobacteri
38	95	7.3	388	16 Q99ZS1	Q99zsl streptococc
39	94.5	7.3	4735	2 O54666	O54666 amycolatops
40	94.5	7.3	9507	2 Q9EWA1	Q9ewal streptomyce
41	94	7.2	530	16 Q8UK38	Q8uk38 agrobacteri
42	92.5	7.1	372	2 Q8RIU8	Q8riu8 corynebacte
43	92.5	7.1	496	10 Q9S833	Q9s833 arabidopsis
44	92	7.1	384	16 P74281	P74281 synechocyst
45	92	7.1	416	16 Q93J85	Q93j85 streptomyce

ALIGNMENTS

RESULT 1  
Q9A6Z1 ID Q9A6Z1 PRELIMINARY; PRT; 261 AA.  
AC Q9A6Z1;  
DT 01-JUN-2001 (TReMBLrel. 17, Created)  
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Transcriptional activator, putative, Baf family.  
GN CC1935.  
OS Caulobacter crescentus.  
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
OC Caulobacter.  
OX NCBI\_TaxID=155892;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 19089 / CB15;  
RX MEDLINE=21173698; PubMed=11259647;  
RA Eisen J., Heidelberg J.F., Laub M.T., Paulsen I.T., Nelson K.E.,  
RA Nierman W.C., Feldblyum T.V., Alley M.R.K., Ohta N., Maddock J.R.,  
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,  
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,  
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,  
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
RT "Complete genome sequence of Caulobacter crescentus.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
DR EMBL; AE005867; AAK23910.1; -.  
DR TIGR; CC1935; -.  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMs; TIGR00671; baf; 1.  
KW Complete proteome.  
SQ SEQUENCE 261 AA; 27965 MW; C19E60D7B0714EF5 CRC64;

Query Match 100.0%; Score 1299; DB 16; Length 261;  
Best Local Similarity 100.0%; Pred. No. 8.4e-89;  
Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MLLAIEQGNTNTMFAIHGASWVAQWRSATESTRTADEYVVVWLSQLLSMOGLGFRAIDAV 60

Db 2 MLLAIEQNTNTMFAIHGASWVAQWRSATSTRTADEYVVWLSQLLSMQGLGFRAIDAV 61  
QY 61 IISVVVPSQIFNLRNLSRRYFNVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAMV 120  
Db 62 IISVVVPSQIFNLRNLSRRYFNVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAMV 121  
QY 121 YPGPLVVIDSGTATTFDVAADGAFEGGIIAPGINLSMQALHEAAKLPRIAIOQPAGNR 180  
Db 122 YPGPLVVIDSGTATTFDVAADGAFEGGIIAPGINLSMQALHEAAKLPRIAIOQPAGNR 181  
QY 181 IVGTDTVSAMQSGVFWGYISLIEGLVARIKAEERGEPMVTVIATGGVASLFEGATDSIDHFD 240  
Db 182 IVGTDTVSAMQSGVFWGYISLIEGLVARIKAEERGEPMVTVIATGGVASLFEGATDSIDHFD 241  
QY 241 SDLTIRGLLEIYRRNTIAES 260  
Db 242 SDLTIRGLLEIYRRNTIAES 261

RESULT 2

QY 9KGH5 PRELIMINARY; PRT; 254 AA.  
AC Q9KGH5;  
DT 01-OCT-2000 (TremBLrel. 15, Created)  
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)  
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)  
DE Hypothetical protein BH0086.  
GN BH0086.  
OS Bacillus halodurans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Bacillaceae; Bacillus.  
OC NCBI\_TaxID=86665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C-125 / JCM 9153;  
RX MEDLINE=20512582; PubMed=11058132;  
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
RA Fuji F., Hiramata C., Nakamura Y., Ogasawara N., Kuhara S.,  
RA Horikoshi K.;  
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
halodurans and genomic sequence comparison with Bacillus subtilis.";  
RL Nucleic Acids Res. 28:4317-4331(2000).  
DR EMBL; AP001507; BAB03805.1; -.  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMs; TIGR00671; baf; 1.  
KW Hypothetical protein; Complete proteome.  
QY SEQUENCE 254 AA; 27907 MW; 0E3B557BA7DAC176 CRC64;

Query Match 46.0%; Score 598; DB 16; Length 254;  
Best Local Similarity 43.9%; Pred. No. 9e-37;  
Matches 112; Conservative 61; Mismatches 80; Indels 2; Gaps 1;  
QY 1 MLLAIEQNTNTMFAIHGASWVAQWRSATSTRTADEYVVWLSQLLSMQGLGFRAIDAV 60  
Db 1 MILVIDVGNNTVLGVYQDETLLVHHWRLATSRQKTEDEYAMTVRSFLDHAGLQFQDIDGI 60  
QY 61 IISVVVPSQIFNLRNLSRRYFNVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAMV 120  
Db 61 VISSVVPMPFSLQCMCKKYFHVTPMIIGPGIKTGLNIKYDNPKEVGADRVNAVAIEL 120  
QY 121 YPGPLVVIDSGTATTFDVAADGAFEGGIIAPGINLSMQALHEAAKLPRIAIOQPAGNR 180  
Db 121 YGYPVAVVDFGTATTYCLINEKKQYAGGVVAPGIMISTEALYHRASKLPRIEIAKP--KQ 178  
QY 181 IVGTDTVSAMQSGVFWGYISLIEGLVARIKAEERGEPMVTVIATGGVASLFEGATDSIDHFD 240  
Db 179 VVGNTIDSQSGIFYGVYVQSDGVVVKRMAQAESEPKVVIATGGVGLAKLIGTSETIDVID 238  
QY 241 SDLTIRGLLEIYRRN 255  
Db 239 SFLTTLKGLQLTYKKN 253

RESULT 3  
QY 9F985 PRELIMINARY; PRT; 258 AA.  
AC Q9F985;  
DT 01-MAR-2001 (TremBLrel. 16, Created)  
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)  
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)  
DE Putative 32 kDa replication protein.  
OS Bacillus stearothermophilus.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Geobacillus.  
OC NCBI\_TaxID=1422;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=V;  
RA Vasquez C., Pichuanes S., Saavedra C.;  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF198621; AAG28531.1; -.  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMs; TIGR00671; baf; 1.  
SQ SEQUENCE 258 AA; 28101 MW; 507B55D695095855 CRC64;

Query Match 43.7%; Score 568; DB 2; Length 258;  
Best Local Similarity 43.5%; Pred. No. 1.5e-34;  
Matches 111; Conservative 54; Mismatches 88; Indels 2; Gaps 1;  
QY 1 MLLAIEQNTNTMFAIHGASWVAQWRSATSTRTADEYVVWLSQLLSMQGLGFRAIDAV 60  
Db 1 MIFVLDVGNNTVLGVYDGDDELKHHWRIETSRGKTEDEYGMTIKALLNHVGLQFSDIDGI 60  
QY 61 IISVVVPSQIFNLRNLSRRYFNVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAMV 120  
Db 61 IISVVVPPIMFALERMLCKYFHKPIIVGPGIKTGLNIKYDNPKEVGADRVNAVAAGIHL 120  
QY 121 YPGPLVVIDSGTATTFDVAADGAFEGGIIAPGINLSMQALHEAAKLPRIAIOQPAGNR 180  
Db 121 YGSPLIIVDFGTATTYCYINEHKQYMGGAIAPIGIMISTEALFARAALPRIEIAKP--DD 178  
QY 181 IVGTDTVSAMQSGVFWGYISLIEGLVARIKAEERGEPMVTVIATGGVASLFEGATDSIDHFD 240  
Db 179 IIGKNTVSAMQAGILYGVQVEGIVSRMKAKSPVPKVIATGGVGLASLIASESNVIDIVD 238  
QY 241 SDLTIRGLLEIYRRN 255  
Db 239 PFLTTLTGLKILYERN 253

RESULT 4

QY 9XHL5 PRELIMINARY; PRT; 259 AA.  
AC Q8XHL5;  
DT 01-MAR-2002 (TremBLrel. 20, Created)  
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)  
DE Hypothetical protein CPE2468.  
GN CPE2468.  
OS Clostridium perfringens.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
OC Clostridiales; Clostridiaceae; Clostridium.  
OC NCBI\_TaxID=1502;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=13 / TYPE A;  
RX PubMed=11792842;  
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,  
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;  
RT "Complete genome sequence of Clostridium perfringens, an anaerobic  
flesh-eater.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).  
DR EMBL; AP003194; BAB82174.1; -.



```
DR InterPro; IPR004619; Baf.
DR InterPro; IPR000515; BPD_transp.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
DR PROSITE; PS00402; BPD_TRANS_INN_MEMBR; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 259 AA; 28819 MW; 0D5FEA3B7A145E10 CRC64;

Query Match 42.4%; Score 551; DB 16; Length 259;
Best Local Similarity 42.7%; Pred. No. 2.8e-33;
Matches 111; Conservative 53; Mismatches 88; Indels 8; Gaps 3;

QY 1 MLLAIEQGNNTMTFAIHDGASWVAQWRSATSTRTADEYVWVLSQLLSMQGLGFRAIDAV 60
Db 1 MILLIDVGNNTNIVLGIHDNEKYIASWRISTDSKTSDEYSIQVMQLFNQAKLPEDVEGI 60

QY 61 IISVVVPQSIFNLNLSRRYFNVEPLVIGENAKLIGDVRIEKPSEAGADRLVNAIGAAMV 120
Db 61 IISVVVPNIMHSLNLMVRKCFCKEPIVVGPGIKTGINIKYDNPKEVGADRIYNAVAAFEK 120

QY 121 YPGPLVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMQALHEAAAKLPRIAIQRPAGNR 180
Db 121 HKKPMIIDIIFGATTFCAITEKGDYLGNGICGIIQISADALFERAAKLPRIELEKP--KS 178

QY 181 IVGTDTVSAMQSGVFWGYISLIEGLVARIKAE-----RGEPTMTVIATGGVASLFEGATDS 235
Db 179 VICKNTVTSMQAGIIYGYIGKVEYIVKRMKKEMMDLGEKEPF-VLATGGLAKLVYSETDV 237

QY 236 IDHFDSDLTIRGLLEIYRRN 255
Db 238 IDEVDRKLTLEGLKILYEKN 257

RESULT 5
Q8R7M2 PRELIMINARY; PRT; 255 AA.
AC Q8R7M2;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Putative transcriptional regulator, homologs of Bvg accessory factor.
GN TTE2381.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
LN [1]
P SEQUENCE FROM N.A.
RC STRAIN-MB4T / JCM11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EMBL; AE013180; AAM25520.1; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 255 AA; 27816 MW; C3C620ECBC8CA6ED CRC64;

Query Match 40.8%; Score 530; DB 16; Length 255;
Best Local Similarity 40.8%; Pred. No. 1e-31;
Matches 104; Conservative 59; Mismatches 90; Indels 2; Gaps 1;

QY 1 MLLAIEQGNNTMTFAIHDGASWVAQWRSATSTRTADEYVWVLSQLLSMQGLGFRAIDAV 60
Db 1 MLLAFDVGNNTNIVMGVFKGKLLHSFRISTDNKNKYDEYGLVNLQINGISLTIEDV 60

QY 61 IISVVVPQSIFNLNLSRRYFNVEPLVIGENAKLIGDVRIEKPSEAGADRLVNAIGAAMV 120
Db 61 IISVVPPMLNTLQVMSLKYFRTKPIVVGPGIKTGINIKYDNPKEVGADRIYNAVAAYEL 120

QY 121 YPGPLVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMQALHEAAAKLPRIAIQRPAGNR 180
```

```
Db 121 YGGPVIVIDFGTATTFCAISEKGEYLGIIAPGLMISADALFQRTAKLPKIDLTQPP--T 178
QY 181 IVGTDTVSAMQSGVFWGYISLIEGLVARIKAEERGEPTMTVIATGGVASLFEGATDSIDHFD 240
Db 179 VINRNTVASMQSGIIYGHVGMVDYIVTRMKGEFAPSAYVYVATGGFANMIAEESKTIDTVN 238
QY 241 SDLTIRGLLEIYRRN 255
Db 239 EMLTLEGLRIIYERN 253

RESULT 6
Q8YAC5 PRELIMINARY; PRT; 259 AA.
AC Q8YAC5;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical protein Imo0221.
GN LMO0221.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
LN [1]
P SEQUENCE FROM N.A.
RC STRAIN-EGD-E / SEROVAR 1/2A;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehlant J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL591974; CAD00748.1; -.
DR MEROPS; M41.009; -.
DR Listlist; LMO00221; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 259 AA; 28187 MW; F106B049D80025B7 CRC64;

Query Match 40.3%; Score 523; DB 16; Length 259;
Best Local Similarity 39.2%; Pred. No. 3.4e-31;
Matches 100; Conservative 59; Mismatches 94; Indels 2; Gaps 1;

QY 1 MLLAIEQGNNTMTFAIHDGASWVAQWRSATSTRTADEYVWVLSQLLSMQGLGFRAIDAV 60
Db 1 MILVIDVGNNTCTVGVEYKQKLLKHWRTTDRHRTSDELGMTVLNFFSYANLTSPDIQGI 60

QY 61 IISVVVPQSIFNLNLSRRYFNVEPLVIGENAKLIGDVRIEKPSEAGADRLVNAIGAAMV 120
Db 61 IISVVPPIMHAMETMCVRYFNIRPLIVGPGIKTGLNLKVDNPREIGSDRIYNAVAASEE 120

QY 121 YPGPLVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMQALHEAAAKLPRIAIQRPAGNR 180
Db 121 YGTPVIVVDFGTATTCYIDESGVYQGGAIAPGIMISTEALYNRAAKLPVDIAE--SSQ 178

QY 181 IVGTDTVSAMQSGVFWGYISLIEGLVARIKAEERGEPTMTVIATGGVASLFEGATDSIDHFD 240
Db 179 IIGKSTVSSMQAGIFYGFVQCCEGIIAEMKKQSNASPPVVYVATGGGLARMITEKSSAVDILD 238
QY 241 SDLTIRGLLEIYRRN 255
```



DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Hypothetical protein lin0253.  
GN LIN0253.  
OS Listeria innocua.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Listeriaceae; Listeria.  
OX NCBI\_TaxID=1642;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CLIP 11262 / SEROVAR 6A;  
RX PubMed=11679669;  
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
RA Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,  
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,  
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,  
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
RA Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,  
RA Madueno E., Maltournam A., Mata Vicente J., Ng E., Nedjari H.,  
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,  
RA Vazquez-Boland J.-A., Voss H., Wehlend J., Cossart P.,  
RT "Comparative genomics of Listeria species.";  
RL Science 294:849-852(2001).  
DR EMBL; AL596164; CAC95486.1; -.  
DR ListiList; LIN00253; -.  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMS; TIGR00671; baf; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 259 AA; 28227 MW; 554B03A0C0EFA64F CRC64;  
  
Query Match 39.7%; Score 516; DB 16; Length 259;  
Best Local Similarity 39.2%; Pred. NO. 1.1e-30;  
Matches 100; Conservative 58; Mismatches 95; Indels 2; Gaps 1;  
  
QY 1 MLLAIEQGNNTMTFAIHDGASWVAQWRSATESTRTADEYVWVLSQLLSMOGLGFRDAIDAV 60  
Db 1 MILVIDGNTNCTGVYKEQKLLRHWRMTDRHRTSDELGMTVLNFFSYANLTSPDIQGI 60  
  
QY 61 IISVVVPQSIFNLNLSRRYFNVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAMV 120  
Db 61 IISVVVPPIHMETMVCVRYFNIRPLVPGIKTGLNKLKVDNPREIGSDRIVNAVAASEE 120  
"Comparative genomics of Listeria species."  
QY 121 YPGPLVVIDSGTATTFDIVAADGAFEGGIAPGINLSMQALHEAAKLPRIAIPAGNR 180  
Db 121 YGTPVIVDFGTATTCYIDEAGVYQGGAIAPGIMISTEALYNRAAKLPKRVDAIE--SSQ 178  
  
JY 181 IVGTDTVSAMQSGVFWGYISLIEGLVARIAERGPMTVIATGGVASLFEATDSIDHFD 240  
Db 179 IIGKSTVAMQAGIFYGFIGQCEGIIAEMKKQSNTSPVVATGGGLARMIKTEKSSAVDILD 238  
  
QY 241 SDLTIRGLLEIYRRN 255  
Db 239 PFLTTLKGLLEIYRRN 253  
  
RESULT 10  
Q8RFE4  
ID Q8RFE4 PRELIMINARY; PRT; 256 AA.  
AC Q8RFE4;  
DT 01-JUN-2002 (TREMBLrel. 21, Created)  
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Bvg accessory factor.  
GN FN0761.  
OS Fusobacterium nucleatum (subsp. nucleatum).  
OC Bacteria; Fusobacteriia; Fusobacterium.  
OX NCBI\_TaxID=76856;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 25586;

RX MEDLINE=21886394; PubMed=11889109;  
RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,  
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,  
RA Vasieva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,  
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,  
RA Fonstein M., Kyrpides N., Overbeek R.,  
RT "Genome sequence and analysis of the oral bacterium Fusobacterium  
RT nucleatum strain ATCC 25586.";  
RL J. Bacteriol. 184:2005-2018(2002).  
DR EMBL; AE010586; AAL94957.1; -.  
KW Complete proteome.  
SQ SEQUENCE 256 AA; 28601 MW; CDF681127F08914B CRC64;  
  
Query Match 32.6%; Score 423; DB 16; Length 256;  
Best Local Similarity 34.2%; Pred. NO. 8.9e-24;  
Matches 90; Conservative 61; Mismatches 96; Indels 16; Gaps 6;  
  
QY 1 MLLAIEQGNNTMTFAIHDGASWVAQWRSATESTRTADEYVWVLSQLLSMOGLGFRDAID 59  
Db 1 MIIGIDIGNTHIVTGIYDNGELISTFRIATNDKMTDEYFYNITKYNEISIKKVD 60  
  
QY 60 VIISVVVPQSIFNLNLSRRYFNVEPLVIGENAKL-----GIDVRIEKPSEAGADRLVN 113  
Db 61 ILISSVVPNIIITFFQFARKYFKVEATIVDLEKKLPFFAKGINY-----TGFGADRIID 115  
  
QY 114 AIGAAVYVPGP-LVVIDSGTATTFDIVAADGAFEGGIAPGINLSMQALHEAAKLPRIA 172  
Db 116 ITEAMQKYPDKNLVIFDFGTATTYDVL-KKGVYIGGGILPGIDMSINALYNGTAKLPVK 174  
  
QY 173 IORPAGNRIVGTDVTSAMQSGVFWGYISLIEGLVARIAERGPMTVIATGGVASLFEA 232  
Db 175 FTFPSS--VLGTDTKQIQAAIFFGYAGQIKHIIKINEELNEEIFVLATGGLGKILSAE 232  
  
QY 233 TDSIDHFDSDLTIRGLLEIYRRN 255  
Db 233 IDEIDEYDANLSKGLTYLYKLN 255  
  
RESULT 11  
Q9RX54  
ID Q9RX54 PRELIMINARY; PRT; 262 AA.  
AC Q9RX54;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Hypothetical protein DR0461.  
GN DR0461.  
OS Deinococcus radiodurans.  
OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;  
OC Deinococcaceae; Deinococcus.  
OX NCBI\_TaxID=1299;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=R1;  
RX MEDLINE=20036896; PubMed=10567266;  
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,  
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
RA Fraser C.M.;  
RT "Genome sequence of the radioresistant bacterium Deinococcus  
RT radiodurans R1.";  
RL Science 286:1571-1577(1999).  
DR EMBL; AF001905; AAF10040.1; -.  
DR TIGR; DR0461;  
DR InterPro; IPR004619; Baf.  
DR InterPro; IPR001220; Lectin\_legB.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMS; TIGR00671; baf; 1.  
DR PROSITE; PS00307; LECTIN\_LEGUME\_BETA; UNKNOWN\_1.  
KW Hypothetical protein; Complete proteome.







```
QY 57 IDAVIISSVVPQSFNLRNLSRRYF-NVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAI 115
Db 58 LTGTAALSTVPSVLHEVRIMLDQYWPSVPHVLIIEPGVRTGIPLLVDNPKVEGADRIVNCL 117
QY 116 GAAMVYPGLVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMQALHEAAAKLPRIAIO 175
Db 118 AAYDRFKAAIVVDFGSSICVDVVSAGGEFLGGAIPGVQVSSDAAARSAAALRRVELAR 177
QY 176 PAGNRIVGTDVTSAMQSGVFWGYISLIEGLVARIKAERG-----EPMVTIATGGVASLFE 230
Db 178 P--RSVVGNKNTVECMQAGAVFGFAGLVGDGLVGRIREDSVGSFVDHDAIVATGHTAPLLL 235
QY 231 GATDSIDHEDSLTIRGLLEIYRN 255
Db 236 PELHTVDHYDQHLTLQGLRLVFERN 260

RESULT 14
-Q9WZY5
D Q9WZY5 PRELIMINARY; PRT; 246 AA.
AC Q9WZY5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein TM0883.
GN TM0883.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
DR EMBL; AE001754; AAD35964.1; -.
DR TIGR; TM0883; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00671; baf; 1.
DR Hypothetical protein; Complete proteome.
W SEQUENCE 246 AA; 27154 MW; 9E0309AD462CF266 CRC64;

Query Match 27.7%; Score 359.5; DB 16; Length 246;
Best Local Similarity 35.5%; Pred. No. 4.4e-19;
Matches 89; Conservative 53; Mismatches 94; Indels 15; Gaps 9;

QY 1 MLLAIEQGNNTMFALHDGASWVAQWRSATSTRTADYVWVLSQLLSMQGLGFRAIDAV 60
Db 1 MYLLVDVGNTHSVFSITEDGKTFRRWRRLSTGVFQTEDELFSLHPL--GDAMREIKGI 57
QY 61 IISVVVPOSIFNLRNLSRRYFNVPEPLVIGENAKLG-IDVRIEKPSEAGADRLVNAIGAAM 119
Db 58 GVASVVPQNTVIERFSQKFIHSPIWV--KAKNGCVKNVKNPSEVGADRVANVAVFK 115
QY 120 VYPGLVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMQALHEAAAKLPRIAIO 179
Db 116 EYKNGIILDMGTATTVDLV-VNGSYEGAILPGFFMMVHSLFRGTAKLPLVEV-KPA-D 172
QY 180 RIVGTDVTSAMQSGVFWGYISLIEGLVARIKAERGEPMTVIATGGVASLFE 238
Db 173 FVVGKDTFENIRLVVNGSVYALEGIIGRIKEVYGD-LPVVLTGGQSKIVK---DMIKHE 228
QY 239 -FDSDLTIRGL 248
Db 229 IFDEDLTIKGV 239
```

```
RESULT 15
O83446 PRELIMINARY; PRT; 273 AA.
AC O83446;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein TP0431.
GN TP0431.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOLS;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
spirochete.";
RL Science 281:375-388(1998).
DR EMBL; AE001220; AAC65417.1; -.
DR TIGR; TP0431; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00671; baf; 1.
DR Hypothetical protein; Complete proteome.
KW SEQUENCE 273 AA; 28472 MW; 439C9C77CB598BC0 CRC64;

Query Match 26.0%; Score 337.5; DB 16; Length 273;
Best Local Similarity 32.9%; Pred. No. 2.2e-17;
Matches 85; Conservative 53; Mismatches 111; Indels 9; Gaps 5;

QY 2 LLAIEQGNNTMFAL--HDGASWVAQ-WRSATSTRTADYVWVLSQLLSMQGLGFRAI 57
Db 1 MLLIDVGNHVVFGIENGGRVRCVRELRLAPDARKTQDEYSLLIHALCERAGVGRASL 60
QY 58 DAVIISSVVPQSFNLRNLSRRYFNVPEPLVIGENAKLGIDVRIEKP--SEAGADRLVNAI 115
Db 61 RDAFISSVVPVLFKTIADAVAQISGVQPVVFGPWAYEHLVPRIPEPVRAEIGTDLVANAV 120
QY 116 GAAMVYPGLVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMQALHEAAAKLPRIAIO 175
Db 121 AAYVHFRSACVVVDCGTALTFTAVDGTGLIQGVVAIAPGLRTAVQSLHTGTQALPLVPLAL 180
QY 176 PAGNRIVGTDVTSAMQSGVFWGYISLIEGLVARIKAERGEPMTVIATGGVASLFE 235
Db 181 P--DSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCCKELGCRCAAVITGGLSRLFSSEVD- 237
QY 236 IDHFDSDLTIRGLLEIYR 253
Db 238 FPPIDAQLTSLGLAHIA 255

Search completed: June 24, 2003, 22:16:14
Job time : 23.8708 secs
```

**THIS PAGE BLANK (USPTO)**